Remote Sensing of Environment 220 (2019) 135–151

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping

Patrick Griffiths[a,b,*], Claas Nendel[c], Patrick Hostert[a,d]

[a] Geography Department, Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[b] European Space Agency, Earth Observation, Science, Applications & Climate Department, Frascati (Roma), Italy
[c] Leibniz Centre for Agricultural Landscape Research, Eberswalder Straße 84, 15374 Müncheberg, Germany
[d] Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

ARTICLE INFO

Keywords:
Crop mapping
Compositing
Time series
Red edge
Land-use
Land cover
Agriculture
Phenology, large-area mapping

ABSTRACT

Many applications that target dynamic land surface processes require a temporal observation frequency that is not easily satisfied using data from a single optical sensor. Sentinel-2 and Landsat provide observations of similar nature and offer the opportunity to combine both data sources to increase time-series temporal frequency at high spatial resolution. Multi-sensor image compositing is one way for performing pixel-level data integration and has many advantages for processing frameworks, especially if analyses over larger areas are targeted. Our compositing approach is optimized for narrow temporal-intervals and allows the derivation of time-series of consistent reflectance composites that capture field level phenologies. We processed more than a year's worth of imagery acquired by Sentinel-2A MSI and Landsat-8 OLI as available from the NASA Harmonized Landsat-Sentinel dataset. We used all data acquired over Germany and integrated observations into composites for three defined temporal intervals (10-day, monthly and seasonal). Our processing approach includes generation of proxy values for OLI in the MSI red edge bands and temporal gap filling on the 10-day time-series. We then derive a national scale crop type and land cover map and compare our results to spatially explicit agricultural reference data available for three federal states and to the results of a recent agricultural census for the entire country. The resulting map successfully captures the crop type distribution across Germany at 30 m resolution and achieves 81% overall accuracy for 12 classes in three states for which reference data was available. The mapping performance for most classes was highest for the 10-day composites and many classes are discriminated with class specific accuracies > 80%. For several crops, such as cereals, maize and rapeseed our mapped acreages compare very well with the official census data with average differences between mapped and census area of 11%, 2% and 3%, respectively. Other classes (grapevine and forest classes) perform slightly less well, likely, because the available reference data does not fully capture the variability of these classes across Germany. The inclusion of the red edge bands slightly improved overall accuracies in all cases and improved class specific accuracies for most crop classes. Similarly, our gap filling procedure led to improved mapping accuracies when compared to nongap-filled 10-day features. Overall, our results demonstrate the valuable potential of approaches that utilize data from Sentinel-2 and Landsat which allows for detailed assessments of agricultural and other land-uses over large areas.

1. Introduction

Most of the pressing challenges humanity is facing today are directly or indirectly related to agricultural production (Johnson et al., 2014; Michael and David, 2017). Given the projected population growth and dietary changes in many of the fastest growing regions in the world, agricultural production needs to be increased and improved to ensure global food and nutrition security. The concept of sustainable intensification describes the avenue towards achieving this goal while acting on land resources and the global climate system with consideration (Tilman et al., 2011; Rockström et al., 2017). In this context, remote sensing for monitoring of agricultural land-use and management plays a key role.

From a remote sensing perspective, agriculture is a complex

* Corresponding author.
E-mail address: Patrick.Griffiths@esa.int (P. Griffiths).

https://doi.org/10.1016/j.rse.2018.10.031
Received 4 October 2017; Received in revised form 18 October 2018; Accepted 20 October 2018

phenomenon which poses unique challenges. For example, the type of crop that is grown on a parcel usually changes within and between years according to the chosen crop rotation. The same crop type can have different temporal and spectral appearance due to local land management, genotype features, site conditions or other environmental factors. Individual fields vary in size by region mainly due to historical land granting and succession practices, sometimes also due to more recent societal approaches to land ownership (e.g. cooperatives in Eastern Europe). They often require high (10–30 m) resolution imagery to be entirely resolved (Ozdogan and Woodcock, 2006; Fritz et al., 2015). Temporal information is usually the key to differentiating individual crop types, making use of unique differences in seasonal growing characteristics and crop phenology. However, detailed spectral information can help distinguishing rather subtle differences among morphologically similar crop types, such as certain cereals (Nidamanuri and Zbell, 2011).

The observational nature of optical, satellite based remote sensing data does not always directly satisfy such requirements. First, despite a fixed orbital repeat cycle, remote sensing observations are often highly irregular across time and space. This is because the effective observation frequency is subject to clear sky conditions but also to other mission related factors (e.g. acquisition schedules, downlink capacity, etc.). The required intra-annual observation frequency for disentangling different crop phenologies can accordingly not easily be obtained for many regions where cloud coverage is frequent. This is exacerbated when national, regional or even global assessments of crop type or condition are targeted, as this requires several orbital imaging swaths to be integrated and analyzed. Coarse resolution (> 100 m) sensors provide frequent temporal observations but lack the spatial detail required to resolve individual fields in many parts of the world (Fritz et al., 2015; Graesser and Ramankutty, 2017), but can work well if field sizes are big as for example in Brazil's agro-industrial landscapes (Maus et al., 2016). Spatially higher resolving sensors acquire data much more infrequently and often additionally lack spectral bands in crucial wavelength domains such as the shortwave infrared or the red edge. While the focus of this present study is on the use of optical remote sensing data, it should be noted that Synthetic Aperture Radar (SAR) data have been successfully used to map crop types (Baghdadi et al., 2009; McNairn et al., 2014; Hütt et al., 2016). The all-weather imaging capacities of SAR sensors represent a key advantage over optical imagery and the high temporal repeat frequency offered by the Sentinel-1 SAR constellation improves has greatly improved the usability for agricultural mapping and monitoring practices (Inglada et al., 2016). Approaches using combinations of SAR and optical data exist, focusing on fusion of the different sensor data (Reiche et al., 2015) or using both sensor observations as independent predictive features (Inglada et al., 2016).

Key agricultural mapping and monitoring applications include crop type, crop management, crop development and health, yield predictions as well as fertilizer or irrigation requirements. For most of these applications, reference data are required that characterize the variability of the targeted phenomena across space and time in a representative manner. This strong dependence on agricultural reference data is unfortunate since such high-quality reference data is often not available or accessible. Consequently, routinely produced, spatially explicit and thematically detailed crop type products are rare. The Cropland Data Layer (CDL) is produced annually since 2008 by the U.S. Department of Agriculture's National Agricultural Statistics Service (Johnson and Mueller, 2010; Boryan et al., 2011). It covers the entire continental U.S. and is based on supervised classification on a per-state level using images acquired by different medium (30–100 m) resolution sensors (e.g. Landsat, Sentinel-2, CBERS, IRS and DMC). The approach makes extensive use of parcel-level information provided by farmers that provides information on crop type per parcel. This data is then used to train supervised classification models and results in very high class-specific accuracies for most of the important crop classes. In Canada, the Annual Crop Inventory provides nation-wide crop maps at the parcel-level based on supervised classification of optical and SAR images from a range of sensors and extensive in-situ data collections since 2012 (Davidson et al., 2017). In Europe, detailed information of agricultural land holdings on the parcel level exists in the form of the land parcel information system (LPIS) and the Geospatial Aid Application (GSAA) data, which tracks individual claims for subsidies made by the farmers (Tóth and Kučas, 2016). Both data sources, LPIS and GSAA represent an integral element for the implementation of the European Union (EU) Common Agricultural Policy (CAP). Unfortunately, LPIS data in Europe is usually treated confidential as it contains sensitive information, for example, on individual land managers and the subsidy payments they obtained. Some European countries such as the Netherlands or Austria have started making LPIS data available, sometimes in the frame of the European Commission (EC) Infrastructure for spatial information in Europe (INSIRES) initiative, while Germany and most other EU countries have not released its data publicly yet. While operational, remote sensing based crop maps at the parcel level over Europe are not available, some first national scale crop maps have been derived from Sentinel-2 and Landsat data in the context of the Sentinel for agriculture project (Sen2Agri, Bontemps et al., 2015). Recent advances in the field of machine learning have considerable potential for advancing data preprocessing, such as sensor fusion and gap filling (Shen et al., 2015; Gao et al., 2006), and to improve the performance of crop- and land-cover classifications through adaptation of techniques such as deep learning (Kussul et al., 2017; Lecun et al., 2015).

Given the pressing societal needs, the unique requirements of agricultural monitoring and the nature of remote sensing observations, analysis and processing strategies are required that allow for accurate and spatially detailed agricultural monitoring over large areas. Fortunately the quality and quantity of medium resolution optical remote sensing data has increased considerably. Landsat data has become freely available in 2008 (Woodcock et al., 2008) and this has drastically increased its usage (Wulder et al., 2012) and sparked many innovative types of analyses (an overview of these is provided in Zhu, 2017). More recently, the European Copernicus program has led to increased volumes of freely available data from Sentinel-2 a/b (S2) and its Multi Spectral Imager (MSI, Drusch et al., 2012). While continuity considerations with Landsat and the Satellite Pour l'Observation de la Terre (SPOT) were important for the mission design, central improvements for MSI include new spectral bands, improved spatial resolution, greater swath width that in twin constellation offers a significantly improved temporal observation frequency.

Despite increased repeat frequencies, optical imagers are still dependent on clear-sky conditions to acquire usable imagery, and persistent cloud coverage can mask out crucial phases of crop development. Integrating data from multiple sensors that share similar observational characteristics can considerably improve the number of clear sky observations (Wulder et al., 2015). Temporal synthesis through image compositing is hence a valuable tool set for combining multi-sensor or multi-mission data. Additionally it is of major relevance for monitoring approaches targeting large areas as it offers a range of advantages for data integration and analysis. First, the pixel-based processing perspective allows exploiting all imagery, including partially cloudy images. Second, temporally and spatially heterogeneous pixel-level observations can be transformed into time series of equidistant and consistent datasets. Such equidistant features are required by many methods (Testa et al., 2018; Udelhoven, 2011) and generally ease the use of more complex processing workflows. Third, compositing also represents an integral step for deriving gridded, higher-level products which are overall still rare for medium resolution data. Compositing can also provide an integrated quality assessment, through consideration of several parameters such as aerosol optical depth or other proxies for atmospheric influences, which can be advantageous for subsequent analyses workflows for example by providing weights for time series fitting (Jönsson and Eklundh, 2004; Maus et al., 2016).

Temporal compositing has only recently become a viable processing option for medium resolution data (Griffiths et al., 2013; Zhu, 2017). Methodologically, there is a range of existing and established compositing approaches. Most approaches follow a best-pixel selection strategy. One of the simplest approaches, maximum Normalized Difference Vegetation Index (max NDVI) compositing selects the observation corresponding to the highest observed NDVI value. The max NDVI approach can be regarded as heritage from early coarse resolution sensors such as AVHRR, as it served the primary purpose of reducing the influence of clouds in the absence of routinely produced cloud masks (Cihlar et al., 1994). The approach has been reused for the Moderate-resolution Imaging Spectroradiometer (MODIS) gridded products by introducing an additional view angle constraint in order to reduce the effect of strongly varying instantaneous fields-of-view due to extreme imaging swath widths (Wolfe et al., 1998; Huete et al., 2002). Other best-pixel compositing approaches select the observations corresponding to the median of a single band or index distribution (Potapov et al., 2012) or use the multi-dimensional median across all spectral bands (Flood, 2013). These approaches have shown to generate very consistent results, but strongly depend on a sufficiently high number of cloud-free observations on the pixel level to work well (Doninck and Tuomisto, 2017). Similarity criteria used to quantify the comparability among available cloud-free observations for a pixel within a given temporal interval have also been used in other best-pixel selection strategies (Frantz et al., 2017; Nelson and Steinwand, 2015). These are, however, conceptually problematic when targeting narrow temporal intervals and intra-annual time series. Operational considerations guided the design of compositing processors for global products, such as WELD, which follows a simple rule set that ensures computational efficiency when processing large data volumes (Roy et al., 2010). Parametric scoring approaches evaluate each pixel observation using a range of different parameters for which a score is produced and the highest (optionally weighted) sum of scores determines the best-pixel selection (Griffiths et al., 2013). The approach was originally proposed allowing for multi-year data integration due to regional data scarcity in the historic Landsat archive, while annual-time series of reflectance composites have been produced following this approach on continental scales in Canada (White et al., 2014), allowing the reconstruction of complex change histories in forest ecosystems (Hermosilla et al., 2016). A recent adaptation of the parametric scoring approach uses local land surface phenology to generate optimized scoring functions for the acquisition day-of-year (DOY), which improves spectral consistency especially in sub-humid environments (Frantz et al., 2017). This algorithm can be fine-tuned according to user requirements by including application specific parameters and regulating their importance by the parameters weights (e.g. maps of aerosol concentrations can be incorporated with a high scoring weight when working in tropical environments typically characterized by high aerosol loadings). Conceptually different approaches calculate new, synthetic values rather than selecting best-pixel values from available Level-2 observations. Examples include weighted averaging (Hagolle et al., 2017), mean value compositing (Vancutsem et al., 2007) or deriving synthetic images from time series fitted harmonic models (Zhu et al., 2015) or by selection of time-series trajectory templates (Vuolo et al., 2017). Summarizing the two main groups of compositing strategies, those that generate new synthetic values, e.g. through averaging, can produce outputs that are very homogenous in appearance, but the values in the composite do represent physical observations. Best pixel selection strategies, on the other hand, strongly depend on the quality of the cloud masking and atmospheric correction to avoid high levels of artifacts.

Here we present an adaptation of the parametric scoring compositing processor (Griffiths et al., 2013) that targets (a) the integration of Sentinel-2 and Landsat reflectance data, (b) the generation of equidistant, dense, and intra-annual composite time series, and (c) provides the basis for a national scale mapping of crop and land cover classes. Our analysis is based on S2A and Landsat-8 (L8) data from 2016 plus the last three months of 2015. We derive a time series of 10-day composites, monthly and seasonal composites. We pose the following research objectives:

- ➢ How do crop and land cover prediction accuracies differ when being based on 10-, 30 or 60-day composite time series?
- ➢ How accurately can crop types be mapped based on the composited time series data and how well does the mapped crop acreage compare to official agricultural census data?

2. Study region Germany

In this study we focus on the Federal Republic of Germany (Fig. 1). The total area is 357,376 km^2. The climate is classified as Cfb according to the Koeppen-Geiger system indicating a mild, marine west coast climate with warm summers and no dry season. Continentality generally increases from West to East. Rainfall and temperature maxima occur during the main summer months, i.e. June–August. Mirroring global trends, the year 2016 tended to be warmer with less precipitation compared to the long term average (DWD, 2016). A pronounced elevational gradient exists from North to South. The Northern lowland is governed by post-glacial deposits, usually forming a gently undulating landscape, which fade to the low mountain ranges (uplands) and finally to proper mountains towards the South, where older geological strata predominate. Soil types in the North typically comprise post-glacial sandy loams with low water holding capacity and more fertile and finer textured soils are found in the South. The Central German Loess belt along the northern foot of the Central Uplands bears the most fertile black soils. Approximately 52% of the territory of Germany is under agricultural use while forests occupy 30% and developed areas comprise about 13%. Approximately 64% of the land under agricultural use is used as cropland (117.6 km^2) while 26% represents permanent grassland (46.9) and the area of perennial crops such as vineyards and fruit tree plantations account for < 2% (Destatis, 2017). Main crops grown are cereals, of which winter wheat (3.13 Mha; 26.3% of cropland), silage maize (2.14 Mha; 18.5%), winter barley (1.27 Mha; 10.8%), winter rye (0.57 Mha; 4.8%), grain maize (0.42 Mha; 3.5%) and triticale (0.40 Mha; 3.4%) are the most abundant. Winter oil-seed rape (1.32 Mha, 11.2%), sugar beet (0.33 Mha, 2.6%) and potato (0.24 Mha; 2.0%) add to the portfolio (Destatis, 2017). However, there are a number of applications conceivable, for which national field-scale information on spatial crop distribution would be essential information to use. These range from national inventories based on simulations to decision-supporting products at farm level, for which a more accurate attribution of crop, soil, weather and production system promises much improved results.

3. Considerations for MSI and OLI integration

Sentinel-2 and Landsat provide overall similar observations due to the continuity considerations in the Sentinel-2 mission design (Table 1). However, there are differences in observation geometry and sensor specification that need to be considered and accounted for in order to allow for a seamless integration. Both sensors provide observations in the main spectral domains with comparable center wavelength and bandwidths. However, Sentinel-2 provides three spectral bands in the red edge domain that lack corresponding measurements in OLI. On the other hand, OLI includes bands in the thermal infrared (TIR) domain, which lack corresponding bands for MSI. Remaining differences in spectral bandwidths and spectral response functions ideally should be adjusted. Spectral bandpass adjustment factors (SBAFs) can be used to further reduce remaining differences. SBAFs for integration of MSI and OLI reflectances can in principle be derived empirically but require simultaneous observations as well as large and representative sampling schemes that allow quantifying the actual difference in spectral

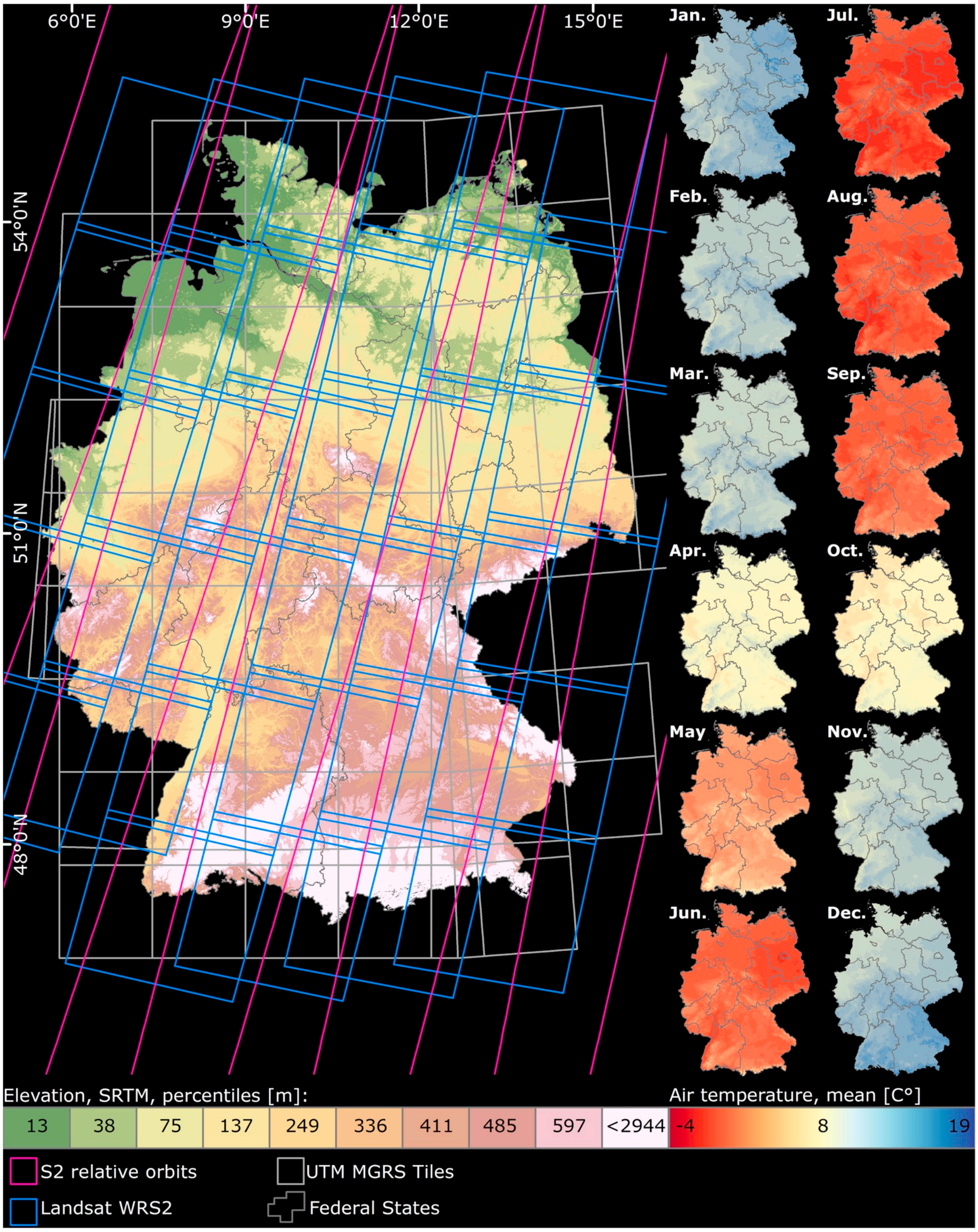

Fig. 1. The study region Germany. The coverage of Sentinel-2 relative orbits, Landsat WRS2 frames and UTM MGRS tiles is provided. Shuttle Radar Topography (SRTM) elevation data is displayed and colors relate to the elevation percentiles in meters for Germany. The monthly mean air temperatures are provided for 2016 (DWD, 2016). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

measurements. Some empirically derived SBAFs have recently been published (Flood, 2017). Atmospherically corrected surface reflectance can be derived directly for both sensors and provides a physically based and largely comparable unit. Ideally the same type of correction and radiative transfer model should be used for both sensors. Several algorithms exist for atmospheric correction of data from both sensors, some providing clear advantages for certain applications and different levels of acceptance and usage in the EO community (Doxani et al., 2018). Bidirectional effects should also be normalized for seamless integration (Nagol et al., 2015; Roy et al., 2016). Sentinel-2 data comes in three spatial resolutions, while the Landsat reflective bands are all processed to 30 m. Varying pixel grids between both sensors need to be aligned and spatial registration uncertainties in multi-temporal acquisitions should be corrected. The level-1 processing of S2 MSI data does not yet make full use of ground control points as the global reference image has not yet been completed (Gascon et al., 2017). The absolute geodetic accuracy of the MSI data is generally assumed to be more precise than that of OLI and subpixel misalignment issues need to be carefully considered (Storey et al., 2016; Yan et al., 2016). Summarizing, both sensor provide similar yet different observations and the methods and tools available to users wanting to combine both data streams do not allow for making these adjustments based on off-the-shelf solutions. Fortunately, we were able to build our research on harmonized Landsat Sentinel-2 data (HLS data, details in the following Section 4.1) that is provided by NASA (Claverie et al., 2018).

Table 1
Specifications of the Sentinel-2 and Landsat-8 missions (FOV = Field of View).

Platform (sensor)	**Sentinel-2 (MSI)**	**Landsat-8 (OLI)**
Swath/FOV	290 km/22°	180 km/15°
Altitude	768 km	705 km
Revisit	10 days (S2A) 5 days (S2A/B)	16 days
Band	**Center wavelength (spatial resolution)**	
Coastal	443 nm (60 m)	443 nm (30 m)
Blue	490 nm (10 m)	482 nm (30 m)
Green	560 nm (10 m)	561 nm (30 m)
Red	665 nm (10 m)	655 nm (30 m)
Red edge	705 nm (20 m)	–
Red edge	740 nm (20 m)	–
Red edge	783 nm (20 m)	–
NIR	842 nm (10 m)	–
NIR	865 nm (20 m)	865 nm (30 m)
Vapor	945 nm (60 m)	–
Cirrus	1375 nm (60 m)	1373 nm (30 m)
SWIR	1610 nm (20 m)	1609 nm (30 m)
SWIR	2190 nm (20 m)	2201 nm (30 m)
TIR	–	10,900 nm (100 m)
TIR	–	12,000 nm (100 m)
PAN	–	590 (15 m)

4. Methods

4.1. Harmonized Landsat-Sentinel data & preprocessing

We used MSI and OLI data from the NASA Harmonized Landsat Sentinel program (HLS, version 1.2, https://hls.gsfc.nasa.gov/). The HLS program (Claverie et al., 2018) provides fully cross calibrated reflectance products for MSI (named M30) and OLI (L30) at 30 m spatial resolution. These products are based on L1C data for MSI and L1T data for OLI. The HLS processing comprises a geometric subpixel alignment of MSI and OLI data to a single MSI tile using an adapted version of the AROP algorithm (Gao et al., 2009). These alignments are made with respect to the baseline version 2.04 processing of S2 Level-1 data with which a revised yaw angle correction was implemented to improve the quality of the geolocation compared to earlier versions. The atmospheric correction of the MSI and OLI data is performed for both sensor data using the LaSRC algorithm (Vermote et al., 2016). Subsequent to atmospheric correction, MSI spectral measurements are adjusted to match those of OLI while bands that lack corresponding counterparts in the OLI (i.e. Red Edge) are not adjusted. For this, a set of band wise empirical SBAFs are applied which have been derived using a global sample of Hyperion spectra. Directional effects are corrected using fixed, MODIS-based coefficients following the approach presented by Roy et al. (2016) resulting in quasi-nadir normalized surface reflectances. Cloud masks complement the HLS image products and are derived using an adapted version of FMASK for MSI and using the LaSRC cloud masking approach for OLI (Claverie et al., 2018). The HLS data is projected to Universal Transverse Mercator (UTM) and is tiled in the Military Grid Reference System (MGRS), which is also used for Level-1 S2 data. The HLS data for Germany covers parts of UTM zones 30 to 33, with each tile extending over 110 × 100 km including an approximately 10 km overlap in X and Y directions (Fig. 1).

We used all data acquired over Germany by Sentinel-2 MSI and Landsat-8 OLI between early October 2015 and the end of 2016. In terms of the regional growing cycle, this period covers the early development of winter crops in 2015 through to the harvesting of summer crops in late summer/autumn 2016. During this period, the S2 constellation was not yet in place, and only S2A was acquiring imagery with a repeat frequency of 10-days over Europe. The HLS data described above contains imagery that corresponds to a total of 220 orbital overpasses by Sentinel-2 and 230 overpasses by Landsat-8. For Sentinel-2, these include 22 overpasses in 2015 and 198 in 2016. For Landsat-8, a total of 47 overpasses occurred in 2015 and 183 in 2016

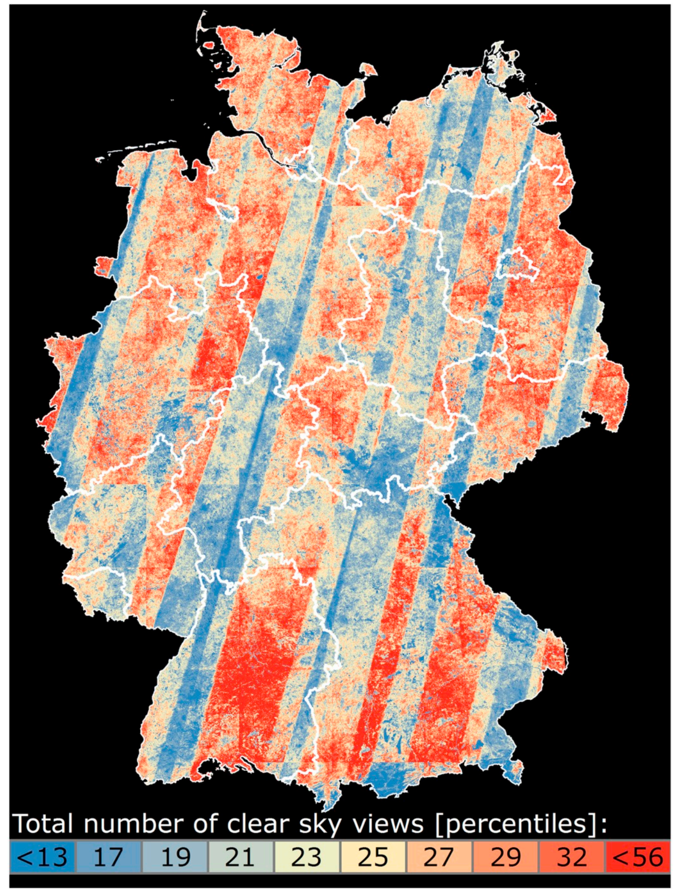

Fig. 2. Number of total clear sky views from the HLS input data per pixel for the time period from DOY 275 in 2015 to DOY 366 in 2016.

(Fig. 2). The highest number of clear-sky views per pixel is 59, the average 23 views. Approximately 25% of the pixels have 19 or less observations for the 15 months study period. High numbers of cloud-free observations generally correspond to areas where adjacent S2 orbits overlap (Fig. 2).

The OLI sensor does not provide corresponding band measurements in the red edge region where MSI provides three spectral bands at 20 m resolution (Table 1). We used a proxy value approach to generate fill values for OLI data at the MSI center wavelength position using band-wise linear interpolation according to Eq. (1) where ρ indicates the reflectance value while λ denotes the center wavelength position for the neighboring S2 spectral bands indicated by j, i and k:

$$\rho_j = (\lambda_j - \lambda_i) * (\rho_k - \rho_i)/(\lambda_k - \lambda_j) + \rho_i \quad (1)$$

The MSI cloud and cloud shadow masks from the HLS data have known shortcomings and higher omission errors compared to the OLI masks presumably at least partly caused by missing TIR measurements for MSI data (Claverie et al., 2018). In order to improve the S2 cloud masks, we applied a number of spectral tests to the input reflectance. We flagged observations as cloud affected if the reflectance in the VIS bands was higher than 20% and the NIR was higher than the SWIR reflectance. Additional tests flagged observations as cloudy if the Haze Optimized Transformation value (see Section 4.2) was greater than zero.

4.2. Compositing

We implemented a best-pixel, weighted parametric scoring processor that was optimized for narrow temporal intervals and multi-sensor data. The optimization for narrow intervals is given though the

Table 2
Parameters used for the parametric scoring compositing processor.

Parameter (abbreviation)	Function type	Parameter weight		
		10-day	MONTHLY	SEASONAL
Distance to Cloud/Cloud Shadow (CDST)	Logistic	1.0	1.0	1.0
Day of Year (DOY)	Gaussian	0.5	0.8	1.0
Sensor (SEN)	Piecewise	0.5	0.5	0.5
Coverage (COV)	Linear	0.25	0.5	0.75
Haze Optimized Transformation (HOT)	Logistic	1.0	1.0	1.0

choice of parameters used and the weights defined for the different scoring functions (Table 2). In general, this type of compositing algorithm generates scores for different parameters for each pixel and each cloud-free observation, applies a user defined weight, sums up the total score and identifies the acquisition with the highest score (Griffiths et al., 2013). The total score per pixel (S_{total}) is thus calculated as the sum of all individual scores (S_i), multiplied with the parameter weight (W_i), divided by the sum of all weights:

$$S_{total} = \sum W_i * S_i / \sum W \quad (2)$$

The observation corresponding to the highest overall score is then identified and all spectral values measured for a given pixel during that acquisition are then transferred into the composite. The relevance of certain parameters for such a compositing approach is different when working with multi-annual (Griffiths et al., 2013; Frantz et al., 2017) compared to annual intervals (White et al., 2014) when targeting a temporally narrow, 10-day interval. We defined a set of five parameters that we consider essential for deriving intra-annual composite time series at narrow temporal intervals, with empirically derived weights reflecting their relative relevance. Table 2 summarizes these parameters, the type of scoring function and the applied weight. For example, the score produced for the acquisition DOY is less relevant on a narrow interval compared to a seasonal or annual interval. It is however still of relevance when determining the best pixel from similarly qualified observations. Thus the assigned weight of 0.5% reflects this lower relevance, as derived from empirical evidence during sensitivity tests.

The highest scoring weight was assigned to Distance to Cloud/Cloud Shadow (CDST) and Haze-optimized Transformation parameters. Both parameters can reduce the susceptibility to errors in categorical cloud masks while not excluding the observations from use in data scarce situations. Pixels observed in close proximity to clouds or cloud shadows are more likely to be affected by remnant cloud cover, haze or fog and cloud related shadowing that was not detected by the cloud/shadow masking algorithm. Users can additionally provide an optional minimum cloud distance parameter (CDST_REQ). Pixels beyond that distance threshold are considered to be reliable observations and obtain a score of 1.0. A logistic scoring function assigns exponentially decreasing scores for pixels with CDST < CDST_REQ. We here set CDST_REQ to 100 pixels.

The Haze-Optimized Transformation (HOT) has been developed to identify haze and thin clouds in Landsat imagery (Zhang et al., 2002) and has been used in the frame of the FMASK cloud masking algorithm (Zhu and Woodcock, 2012) and more recently also in Landsat compositing approaches (Lück and van Niekerk, 2016; Frantz et al., 2017). We derived the HOT according to (Zhu and Woodcock, 2012):

$$\text{HOT} = (\rho_{blue} - 0.5 * \rho_{red} - 0.08) \quad (3)$$

Pixels with higher values indicate a higher likelihood to be affected by haze or thin clouds. The logistic scoring function for the HOT score is calculated as:

$$S_{HOT} = 1 / \left(1 + \exp\left(\frac{10}{0.02} * (HOT + 0.075)\right)\right) \quad (4)$$

The calculation of the DOY score is based on a Gaussian scoring function, where μ indicates the target DOY, σ is the standard deviation and x_i the DOY for which the score is calculated:

$$S_{DOY} = \frac{1}{\sigma\sqrt{2\pi}} * \exp\left(-0.5\left(\frac{x_i - \mu}{\sigma}\right)^2\right) \quad (5)$$

The score is scaled between 0 and 1 by dividing by the maximal possible score. The standard deviation σ is set to a value of 2.4 for the 10-day interval, 5 for the monthly and 12 for the seasonal composites. This ensures sufficient prioritization of center dates while low values are assigned to the edges of a given temporal interval. The sensor score provides a simple mean for prioritizing observations made by one sensor over those of another sensor. In the frame of this study, we favor observations made by Sentinel-2 over those made by Landsat-8 mainly due to the fact that no observations are made in the red edge bands for OLI and that we use proxy values derived through linear spectral interpolation. Thus, a score of 1.0 is assigned to pixels observed by Sentinel-2 while observations made by Landsat-8 obtain a score of 0.8, which appeared to create the most homogeneous results in initial tests. The coverage score COV is introduced to favor tiles that provide cloud-free coverage for a large area. This score is included in order to enhance synoptic coverage of a few acquisitions and prevent compositing data for unnecessary man acquisitions. The COV score is identical to the percentage of cloud-free coverage of the area of interest i.e. the currently processed tile.

We first parameterized the compositing algorithm to generate a time series of 10-day interval composites, resulting in a total of 45 reflectance composites for 9 spectral bands across the 62 MGRS tiles (Fig. 1). We consider this the benchmark composite dataset for this study. Besides the best-observation reflectance composite, metadata attributes are stored in a separate metadata raster file. The values stored per pixel include the acquisition DOY, the observing sensor, the weighted sum of scores and the number of cloud-free observations in a given temporal interval. Next we generated a series of monthly composites using the monthly calendar days as intervals and ranging from October 2015 to December 2016. Finally, the processor was parameterized to generate seasonal composites (i.e. one composite for each calendar-based season) using the following DOY ranges as temporal intervals:

- Spring: 095–155, target DOY 125 (04 May)
- Summer: 189–249, target DOY 219 (06 August)
- Fall: 280–340, target DOY 310 (05 November)
- Winter: 004–064, target DOY 034 (03 February)

These ranges are centered on the meteorological seasonal median and include 30 days before and after that date. Please note that both the monthly and seasonal composites are generated from the original input data and not derived from the 10-day composites. The code is implemented in Python and makes extensive use of the GDAL, NUMPY and SCIPY libraries as well as SCIKIT-LEARN (Pedregosa et al., 2011) for the subsequent machine learning analysis and mapping. The RIOS package is used for tile processing and parallelization (Clewley et al., 2014). The entire compositing workflow including all of the scoring functions is based on array operations on multi-dimensional NUMPY arrays ensuring overall computational efficiency. The implementation also makes extensive use of the lazy-evaluation concept (Gorelick et al., 2017) and the GDAL virtual raster format. This means that nearly all intermediate steps and temporary rasters are stored as XML representations of the data and processing steps are only evaluated on a per-pixel basis once the final results are written to disk.

4.3. Temporal gap filling

Subsequent to compositing, the benchmark 10-day interval composite time series underwent a temporal gap-filling processing stage. The benchmark dataset represents a quality-screened series of temporally binned observations at the narrow 10-day interval and this allows for different temporal gap filling approaches. We implemented a simple and computationally efficient solution based on band-wise linear interpolation, also acknowledging that potentially long gaps due to cloudiness over Germany do not allow for more sophisticated gap-filling. Arrays were generated that record the preceding and subsequent valid observations spectral value as well as DOY taken from the metadata raster (i.e. actual observation DOY, not mean interval DOY) to calculate an interpolated reflectance value for a given band and interval. The calculation of an interpolated reflectance value ρ for a given band and temporal interval j is then performed using array arithmetics and the DOYs of the preceding and subsequent observations t_i and t_k:

$$\rho_j = \frac{(t_j - t_i)*(\rho_k - \rho_i)}{(t_k\_t_i)} + \rho_i \tag{6}$$

The result of the interpolation procedure is becoming rather unrelated to the preceding and following observations when the interpolation fills wide gaps of 10-day intervals. We therefore introduced a *max_interval* parameter that limits the number of intervals over which the interpolation can be performed. Empirical tests here suggested a *max_interval* of 10 (i.e. maximum of 100 days) being the maximum limitation while allowing for achieving full coverages throughout the interval time series.

4.4. Reference data

A key dataset for this study was the GSAA data from the LPIS (in the context of this study we simply refer to this reference data as LPIS) that was available for 2016 covering three German states: Mecklenburg-West Pomerania (German: Mecklenburg-Vorpommern - MV), Brandenburg (BB) and Bavaria (German: Bayern - BY). The LPIS data originates from the Integrated Administration and Control System (IACS) of the EU and is the basis for the payment of agricultural subsidies in the frame of the EU's CAP. The dataset represents self-reporting data, i.e. land owners use an online Geographic Information System (GIS) to digitize their parcels on orthophotos or very high resolution satellite imagery. The level of geometric accuracy is usually very high, so that all non-farmed landscape elements in a parcel (e.g. hedgerows or windmills) are excluded from the digitized areas. While overall the LPIS data can be considered very reliable, it can contain errors e.g. due to false claims or digitization errors. The LPIS provides e.g. the crop type or other agricultural land-use for a given parcel representative for months of June to August. Not all agricultural parcels are part of the LPIS because for example some farmers might not apply for any subsidies or a parcel does not qualify for payments in the frame of the CAP. LPIS accordingly does neither provide a wall-to-wall map of all agricultural plots in a country, nor does it cover other land covers or uses than those relating to agriculture. The statistical distribution of parcel sizes in the LPIS data showed that the highest number of parcels was available for BY (> 1 M), where also the highest share of small parcels (< 1 ha) was found (47%). In BB and MV only 24% and 30% or parcels were smaller than 1 ha, and the share of parcels larger than 10 ha accounted for 24% and 28%, respectively, while only 1.3% of parcels were > 10 ha in BY.

For our crop classification legend, we included all classes that accounted for at least 1% of the LPIS area in each of the three federal states (Table 3). We included some additional crop classes that accounted for < 1% as these are of specific relevance in other areas of Germany. These include, for example, sugar beet and grapevine. For some agricultural land-use classes the LPIS provides a taxonomy that is not directly usable in a remote sensing analysis. This especially concerns the different grassland classes (Section 4.5).

Table 3
The main agricultural classes in the LPIS for the states of Mecklenburg-West Pomerania (MV), Brandenburg (BB) and Bavaria (BY). Classes that covered at least 1% of the LPIS area were included.

Class	Percentage of LPIS area			Number of parcels		
	MV	BB	BY	MV	BB	BY
Meadows	4.5	5.2	24.9	12,485	17,936	645,716
Winter wheat	23.3	12.5	18.5	12,837	9602	198,838
Maize (silage)	10.3	11.8	13.6	8553	10,630	174,705
Mowed pastures	12.0	15.8	8.6	24,963	37,008	114,757
Winter rapeseed	17.1	10.1	3.3	9019	6761	40,289
Winter rye	4.1	13.4	1.2	4059	15,475	17,504
Winter barley	9.1	6.9	8.3	5040	5381	103,441
Maize (other)	< 1.0	1.3	3.9	–	1329	47,666
Winter triticale	< 1.0	3.1	2.4	–	3691	34,950
Pastures	2.7	< 1%	1.3	6470	–	28,861
Fallow	2.1	2.4	< 1.0	15,459	11,331	–
Spring barley	1.0	< 1.0	3.1	1170	–	42,189
Planted grasses	1.6	2.7	< 1.0	3666	9965	–
Clover grass	< 1.0	< 1.0	2.0	–	–	37,665
Sugar beet	1.8	< 1.0	< 1.0	931	–	–
Maize (energy)	< 1.0	1.9	< 1.0	–	1672	–
Mountain meadows	–	–	1.2	–	–	5486
Alfalfa	< 1.0	1.1	< 1.0	–	1463	–
Spring oat	< 1.0	1.1	< 1.0	–	2534	–
Winter spelt	< 1.0	< 1.0	1.0	–	–	11,792
Lupine	< 1.0	1.0	< 1.0	–	1477	–
Durum wheat	1.3	< 1.0	< 1.0	870	–	–

We used an additional reference dataset for training and validation of broad forest classes as well as a built-up and water classes. We used data from the Land-use/Cover Area frame Survey (LUCAS, Palmieri et al., 2011; Karydas et al., 2015) sample of the European Commission's statistics office Eurostat. The LUCAS survey is carried out every three years and is defined by a regular grid of sampling points at two kilometer distance and covering the entire territory of the EU. Each sample is first photo interpreted and then a subsample of points is visited in the field where land cover and land-use types are registered. Additional attributes on the type of observation (e.g. on the point, from a certain distance) or the shape and size of the mapped landscape element (e.g. size of parcel, width of linear element) are collected during the field visits. As of today, the last LUCAS survey was carried out in 2015 and we utilized this data for training and validation of the forest, built-up and water classes. In general these classes are relatively stable in Germany and we do not expect many samples of these classes being affected by land changes between 2015 and 2016. We applied a conservative filtering of the sample points based on the attributes describing the sample interpretation (e.g. only directly observed points, no narrow linear features, etc.). After filtering, the LUCAS survey data did not provide a sufficient number of points that would allow for a balanced sample across all classes for training and validation. We therefore sampled additional points from other datasets. For the forest classes, we used orthophoto interpreted biotope maps that were available for the state of Brandenburg. For the built-up and water classes, we used the Copernicus high resolution layers obtained from (Gallego et al., 2016), which we resampled and reprojected to the 30 m UTM grid.

4.5. Training & validation data

The final class legend used for mapping consisted of 12 classes (Table 4). Some classes, such as grassland or maize, contain a number of subclasses that can vary according to the LPIS data from the different states. For grassland we combined subclasses such as meadows, mowed pastures, and planted grasses from the LPIS data from all three states

Table 4
Final legend of target classes used for mapping of crop types and land cover in Germany, including main characteristics, such as the main time window for sowing, time of peak greenness and the main harvesting time window. Note: cultivar choice, designated product use and the climatic gradients across Germany produce considerably large time windows for some crops.

Class	Sowing time window	Peak greeness	Harvesting time window
Grassland	Perennial	Mid April–Mid July	June–Mid September
Winter cereal	Mid September–Late October	May	Mid June–Early August
Maize	Late April/Early May	Mid June–Early August	Mid September–November
Winter rapeseed	Mid August	Late March–Late April	July
Spring cereals	Mid March	June	July–Mid August
Sugar beet	Mid April–Early May	July–August	Late September–Early November
Potato	Mid April–Mid May	June–July	Early July–Late September
Grapevine	Perennial	June–August	Mid August–Mid October
Deciduous & mixed forest	Perennial	June–August	–
Coniferous forest	Perennial	–	–
Built-up	–	–	–
Water	–	–	–

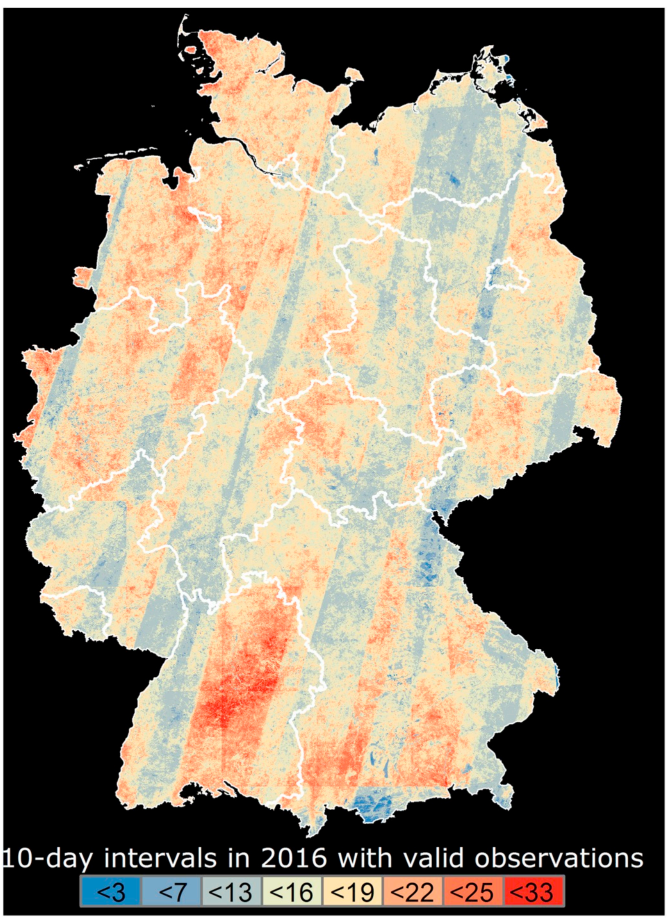

Fig. 3. Number of multi-sensor composites at 10-day interval in 2016 with non-interpolated observations.

and included mountain meadows only for Bavaria as this class does not exist in the other LPIS data. Land management on all of these classes includes mowing, while – depending on the class and federal state - grazing by livestock can additionally occur. Thus, these classes cannot be readily separated based on their spectral and temporal characteristics and we used one broad grassland class that contained several grassland subclasses. For all LPIS classes, we discarded all parcels that were smaller than 1 ha in size. We then restricted the individual parcel area by using a 30 m inside buffer and then split the parcels per class and state into 70% parcels for training and 30% for validation. We additionally restricted samples for training to areas with a sufficient number of valid (i.e. cloud-free, not interpolated) observations. For this, we calculated the number of valid observations in 2016 (Fig. 3) and tested different thresholds. The average value was 15 and we finally used a value of 20, as higher values limited the reference data too restrictively. The threshold of at least 20 observations within the 10-day composites was determined empirically and seemed to provide a good balance between excluding too much area while achieving satisfying results for the class predictions. We used a final sample size of 1500 and 1000 points for training and validation, respectively, for each of the 12 target classes in Table 4. For each state and target class we sampled pixels from the LPIS proportionally according to the relative contribution of a subclass to the target class and potentially allowing for several pixels being selected for a parcel. For example, we combined samples from three subclasses for maize (maize for bio-energy, silage maize and grain maize) according to how much of the maize area each of those classes' subclasses contributed for a given federal state.

4.6. Agricultural census data

On the foundation of EU Regulation No 1166/2008 on farm structure surveys and the survey on agricultural production methods, EU member states conduct national agricultural censuses every couple of years. The most recent census launched by the Federal Statistics Office of Germany surveyed German agriculture in 2016 (Destatis, 2017). The census is based on questionnaires that are sent out to all farmers in Germany. The results provide for each German federal state, among others, the total area that was used to grow a certain type of crop. We used the census results to evaluate how realistic the mapped area per crop type and federal state are. While this does not qualify as a full validation, it provides a valuable addition to the point-based validation.

4.7. Classification

We used a Random Forest (RF) classifier (Breiman, 2001) based on the implementation available from the Scikit-learn library to map our classes (Pedregosa et al., 2011). Random Forests generate a multitude of decision trees by randomly drawing samples with replacement from the training data and determining the best split at each decision tree node by considering a maximum number of randomly selected features (max features). We tested different parameter ranges for the number of trees and for max features and finally used 1000 trees and 10% of the input features considered at each split to parameterize the RF classification models. We trained each RF model for the 10-day, monthly and seasonal composites.

The spectral features we uses for parameterizing the RF models and predicting the maps were the 2016 time series of multi-spectral values, totaling 324, 108 and 36 spectral bands for the 10-day, monthly and seasonal composites (Table 1). The inclusion of the 2015 composites covers a crucial period of phenological development for the winter

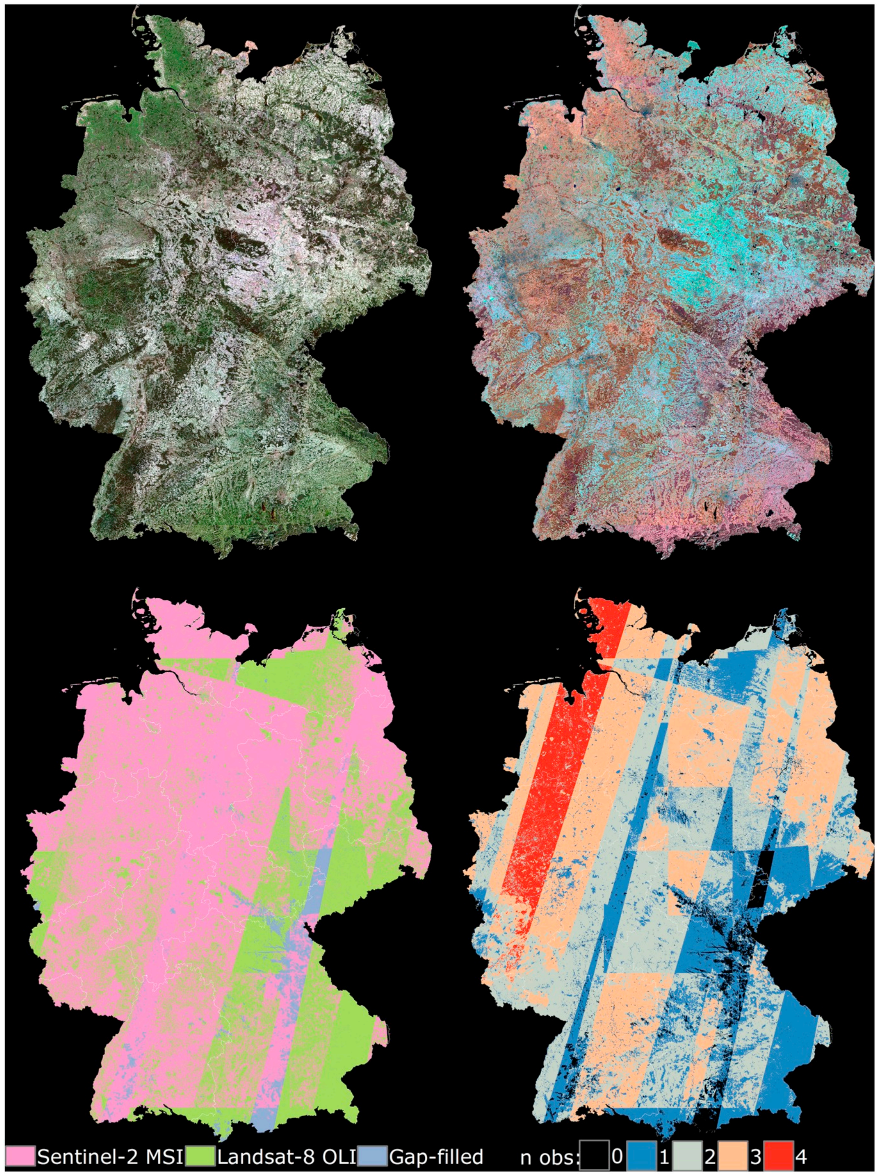

Fig. 4. Example output of the 10-day interval composites for the target DOY 255 in 2016. The top row shows a true color (left, RGB = red, green, blue) and false color (right, RGB = NIR, SWIR1, red edge 1) display. The second row shows the corresponding sensor flag (left) and the number of clear sky observations for the interval period (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

crops and can potentially lead to improved discrimination against other classes. On the other hand, due to different crop rotations, the use of the 2015 composites for mapping can also introduce confusion as the same 2016 crop could have different preceding crops or cover types.

In order to see if the inclusion of the red edge bands to the classification models leads to an improvement in mapping accuracy, we parameterized separate models where we omit the three red edge bands (compare Table 1). For the 10-day composites, we run two additional tests for the RF models parameterization: (1) using the non-gap filled composite time series, and (2) using the non-gap filled 10-day composites while excluding the red edge bands. Additionally we derived variable importance metrics per target class and spectral band. For this, we parameterized several RF models using a single spectral band from all 10-day interval composites in 2016. In each model, we then trained the RF as one class against all other classes using the multi-temporal features per band.

5. Results

5.1. Composite time series

Our compositing algorithm generated a time series of 45 reflectance composites with 9 spectral bands for the benchmark 10-day interval composite run, extending from October 2015 to December 2016. An example for target DOY 255 in 2016 (corresponding to September 11th) is provided in Fig. 4. The 10-day compositing interval ranged from DOY 250 to 259 and cloud free sensor observations were available for over 96% of pixels (Fig. 4, lower right). For the majority of pixels (37%) two clear sky observations were available in this period and about a quarter of the pixels had either one or three observations available. Less than 7% of the pixels had four clear sky views during the 10-day interval. 61% of these pixels were observed by S2 MSI, while 35% were observed by L8 OLI and 4% of pixels obtained values through our temporal gap filling procedure (Fig. 4, lower row). At the scale provided, the result is visually homogenous and free of any apparent artifacts.

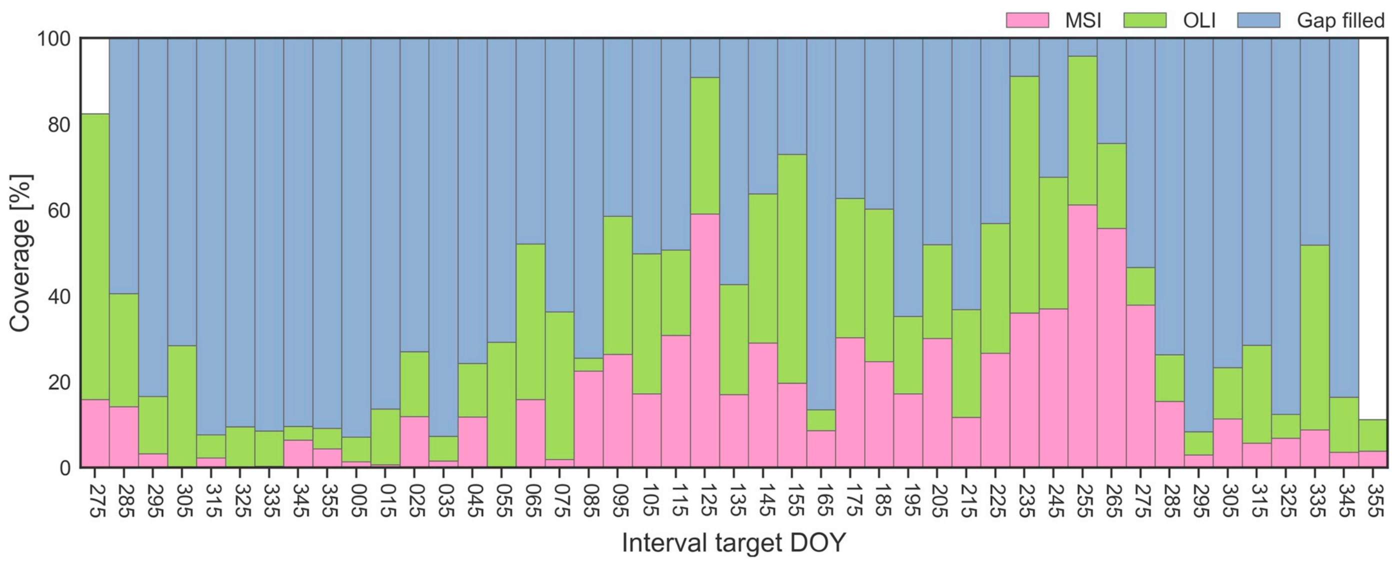

Fig. 5. Coverage of Germany by sensor for 10-day multi-sensor composites from target DOY 275 in 2015 to 355 in 2016.

The coverage with clear sky observations was not always as good as for DOY 255 in 2016. The composite for target DOY 275 in 2015 achieves over 80% of coverage from approximately 60% OLI observations and < 20% contribution from MSI acquisitions (Fig. 5). The following period from DOY 285 in 2015 to DOY 65 in 2016 corresponds to the winter months and observation density is generally very low. For example, the 10-day composite for target DOY 35 in 2016 had only cloud-free observations for 7.2% of the pixels considering both sensors (OLI contributed 5.7%, MSI 1.5%). With a few exceptions, the intervals during the summer months allowed filling at least 40% of the pixels with cloud-free observations during 10-day periods from either of the two sensors. The share of observations acquired by S2 increased over the course of 2016, but the contribution of L8 OLI remained substantial (Fig. 5). The temporal gap filling allowed achieving 100% coverage for all 10-day interval composites using a max_interval value of 100 days. Fortunately the first composite for DOY 275 in 2016 had > 80% of sensor observations which aided gap filling for the composites during the subsequent winter period.

Fig. 6 shows three examples of single-pixel temporal trajectories from the training data. All three examples feature some early OLI observations during autumn 2015 after which no observations were available during the winter months. Clearly, the period where the different spectral bands show the greatest variability is the period of main crop development during DOY 95 to 275 in 2016 that also coincides with the period with the least amount of gap-filled values. The depicted (second) red edge band at 905 nm during some periods shows some behavior that is largely uncorrelated to the RED or NIR bands, for example during DOY 175 for the potato example and during DOY 85 in the maize trajectory. Visualized at this scale some pixel level artifacts are discernible. The temporal interpolation provides fill values for these unobserved periods, but do not reconstruct non-linear changes on individual parcels not captured by any satellite overpass. The band-wise linear interpolation may still produce non-linear gap-fills for unobserved 10-day intervals, as the visible red and the near infrared can of course behave differently over time. The main period of crop development in 2016 is in all three examples observed by OLI or MSI starting around DOY 65 (March 5th). For both, potato and sugar beet, the main green-up occurs around DOY 145, while the growing season for maize starts shortly after. The quality of the composites suffers from occasional cloud or cloud shadow remnants, as for example at DOYs 95 and 225 in the potato profile (Fig. 6a). The maize trajectory is negatively affected by shortcomings of cloud-shadow masking that occurred during DOY 165 in 2016. A similar effect can also be observed for sugar beet at DOY 185 in 2016. The 2016 potato harvest is discernable around DOY 235 when the SWIR trajectory intersects the NIR reflectance (occurring possibly slightly earlier, as DOY 225 is affected by cloud shadows). Maize harvest occurs around DOY 255 in 2016, while the sugar beet harvest occurs later but is not clearly discernable due to no available observations during DOY 285 to 305.

The time series profiles extracted from the training data clearly show how the winter versus spring or summer crops follow similar temporal reflectance patterns (Supplementary material, S02–S12). Accordingly, the reflectance values in the VIS and SWIR bands decrease when the canopy consolidates while the reflectance in NIR and red edge bands behaves conversely. The built-up and water classes interestingly follow a seasonal patterns while the variability (one standard deviation is indicated by the grey patterns) is much larger as for example for the forest classes.

The variable importance suggests that clearly different spectral features are of varying relevance during different parts of the year. For example S-14 shows how the relevance of all bands (except SWIR) for the discrimination of rapeseed is high during the DOY intervals 125 and 135. As indicated by the phenology time series plots (S-07) this is the period where high NIR and red edge reflectance occurs, due to the bloom and canopy closure during spring. For the discrimination of potato (S-17) the red edge bands also have high variable importance scores during the main growth period (S-05).

5.2. Crop type and land cover mapping

A crop and land cover map was obtained for the entire country by predicting the land-cover class for each pixel based on the full vector of spectral values obtained from the gap filled 10-day interval composite time series for 2016 (Fig. 7).

Spatial patterns of land-use and cover are well preserved in the map. The main centers of crop production in Germany (the North Sea polders, the North-Eastern ground moraines, the Central German Loess belt, the Loess areas of the South German Scarplands and the flood plains along the main rivers) become apparent in shades of blue and magenta, indicating the prevalence of winter and summer cereals (Fig. 7, Frame 3) or in hues of yellow and orange if crop rotations include a greater abundance of rapeseed (North-East and West), potato (Central North) or maize (several regions). The crop land as characterized here mirror the spatial distribution of high to medium agricultural yield potentials as identified by Soil Quality Rating (Mueller et al., 2012; BGR, 2014). The main grassland areas are identified in the

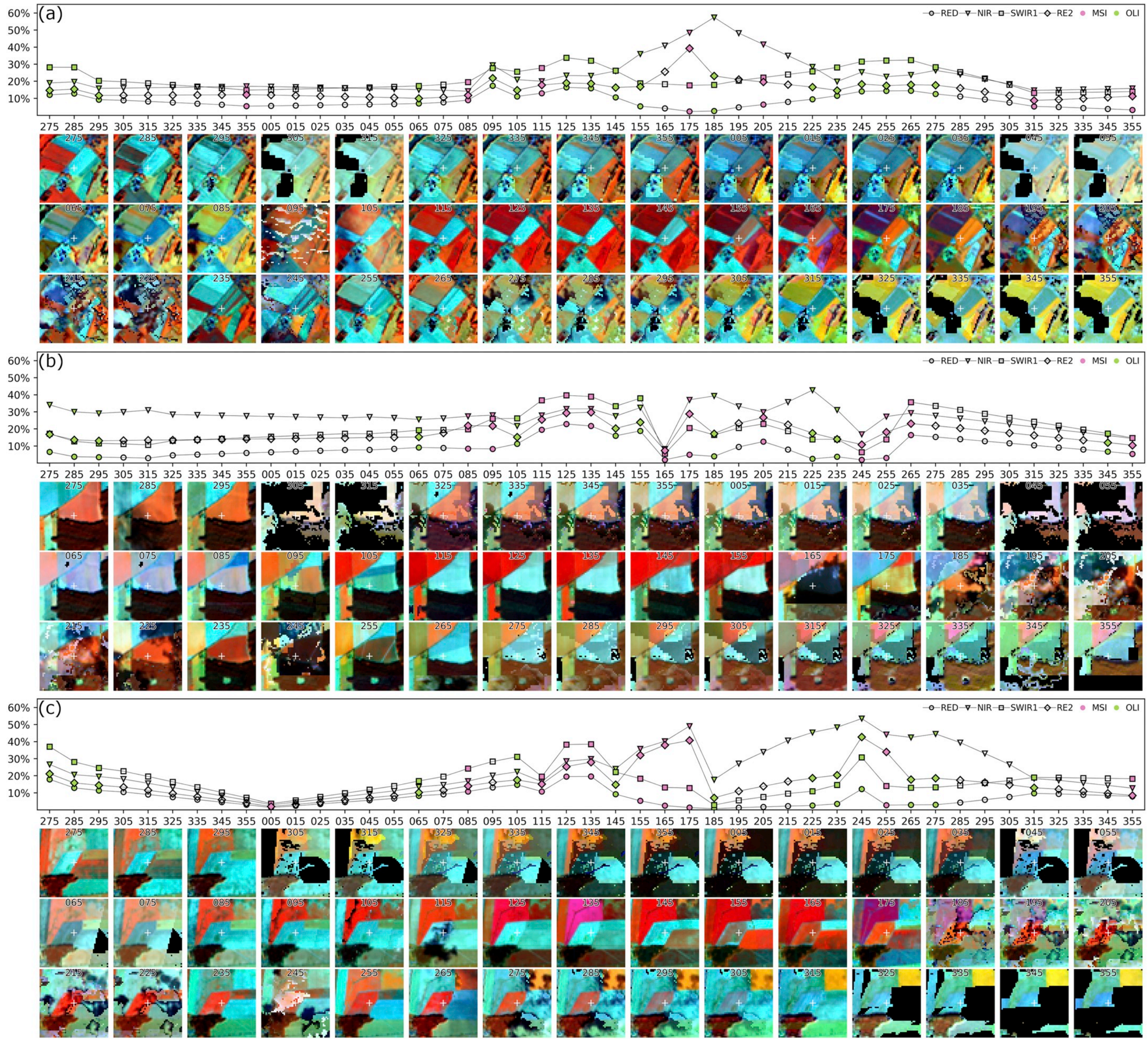

Fig. 6. Examples of time series profiles taken from the 10-day interval composite data set and spanning the period from DOY 275 in 2015 to 355 in 2016. Example (a) is potato, (b) is maize and (c) is sugar beet. Each point in the trajectory corresponds to a 10-day interval. Interpolated (i.e. gap filled) values are shown as unfilled symbols, while observations by S2 MSI are in magenta and those observed by L8 OLI are in light green. The visualized trajectories correspond to the NIR, SWIR (1600 nm, compare Table 1), Red and Red Edge (740 nm) bands. The corresponding image chips for each interval are shown below (RGB = NIR, SWIR1, Red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

North-West (Schleswig-Holstein and Lower Saxony) and far South (Southern Bavaria), mostly located on poor, shallow or not well-drained soils (often shallow groundwater tables) or on slopes too steep for highly mechanized agriculture. At this stage, we refrained from separating different grassland types (pastures and meadows, managed at different intensities), but the presented approach bears the potential to also look more closely into these.

5.3. Accuracy assessment

The overall accuracy (OAC) of the crop and land cover map using the 10-day interval composites features is 80.92% for the 12 classes (Fig. 8). The OAC increased to 81.37% after applying a 5-pixel minimum mapping unit. The OAC obtained using monthly features is just slightly below (79%). Using the seasonal features resulted in lower OAC of 74.6%. In all cases, the OAC achieved when including the red edge bands was higher than if the red edge bands were left out (Fig. 8). This difference was rather gradual for the 10-day composites (0.2%), but slightly higher when using monthly (0.4) or seasonal features (1.0%). When mapping was based on the non-gap filled 10-day composite time series, the OACs was reduced by 0.9% (80.0%) and when we additionally excluded the red edge bands from the non-gap filled time series the OAC was reduced by an additional 1.5% (79.4%).

Class-specific user's (UAC) and producer's (PAC) accuracies for the crop classes are, with few exceptions, highest for the 10-day features and lower for monthly and lowest for the seasonal features. The most pronounced differences in class specific accuracies between different input feature sets are observed for both cereal classes, potatoes, maize

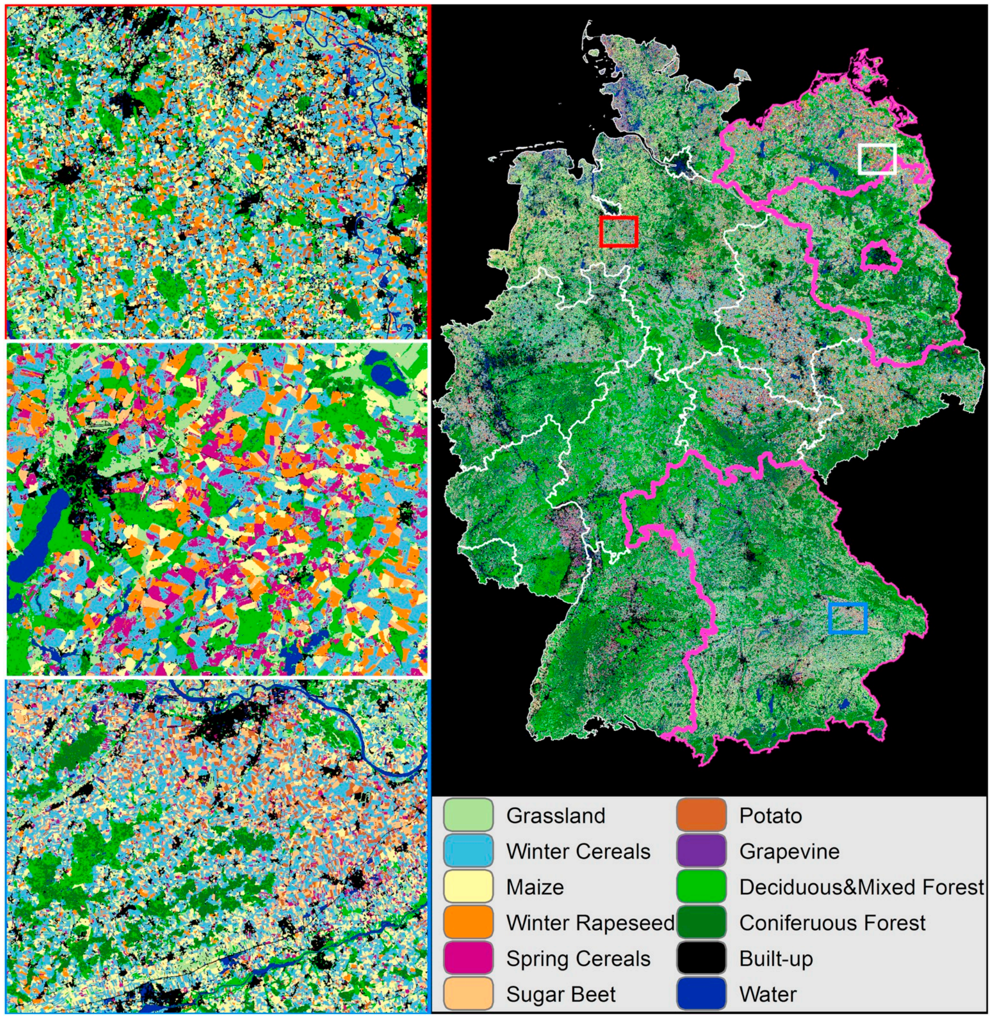

Fig. 7. Result of the wall-to-wall crop type mapping using the benchmark 10-day interval composite time series. The state borders are shown as white lines, while the three states for which reference data was available are shown with magenta outlines (from North to South: Mecklenburg-Vorpommern (MV), Brandenburg (BB) and Bayern (BY)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and sugar beet. Less pronounced are the differences for the grassland class, which achieved 79.5% PAC and 72% UAC. The non-agricultural classes, i.e. the rather static forest, built-up and water classes show only minor differences in class-specific accuracies for the different input feature sets. However, there is some degree of confusion between the built-up and water classes with for example winter cereal. The seasonal variability of spectral values for the training data of these classes (Figs. S1–S10) is pronounced suggesting a certain degrees of mixtures of materials so that spectral temporal similarities result over the course of the time series. The overall best-performing crop classes are rapeseed, sugar beet and maize with PAC and UAC values above 80% with the exception of maize, which is associated with a slightly higher commission error. Winter cereal (> 80% PAC and UAC) is mapped with higher accuracy than spring cereal using the 10-day interval composites. Potato and Grapevine classes achieved rather low PAC values (49% and 62%, respectively). When excluding the red edge bands from the features used in the classifications, class specific accuracies showed different responses. For the grassland class, incorporating the red edge bands led to slight decreases in PAC and UAC values when using 10-day, monthly and seasonal features. Conversely, maize class specific accuracies were improved in all cases when the red edge bands were included with most differences being around 2%. In general, the differences related to red edge bands are were largest in the seasonal and most gradual in the 10-day time-series. Moreover, crop classes showed the greatest effect when using or omitting the red edge bands.

When mapping was based on nongap-filled 10-day features most class specific accuracies achieved lower values. For example, the PACs were 2% lower for winter cereal and maize and 3.4% lower for potatoes. The drop in UACs was at least 4% for maize, summer cereal and potato. The only exception was sugar beet where the UAC and was 9% higher when nongap filled features were used. All nonagricultural classes decreased by between 2% and 4% in UAC. These class specific accuracies were further reduced when we additionally omitted the red edge bands from the nongap filled 10-day composites.

Moderate confusion occurs between the winter and spring cereal classes (Table 5). The before-mentioned high commission error for maize is partly based on confusion with potato. Also noteworthy is, for example, the confusion between grassland and grapevine as well as spring cereal and potato. For the nonagriculture classes, most confusion exists between both forest classes between built-up and water.

5.4. Comparison with the agricultural census data

Fig. 9 shows scatterplots comparing the mapped area against the area reported in the 2016 agricultural census for different crop types aggregated to the level of the federal States. The estimate of the mapped grassland area compared relatively well with the census data with a tendency towards an overestimation of the mapped area. The average difference between mapped and census area over all states (excluding the city states of BE, HH and HB), for the grassland class was 2%. A

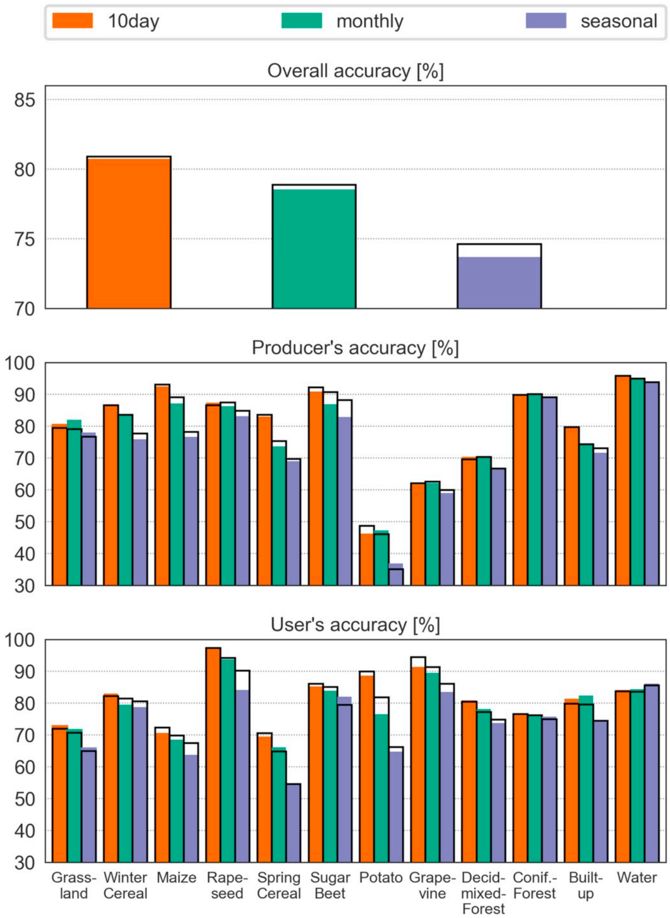

Fig. 8. Validation results of the classifications using only the 2016 10-day, monthly and seasonal composites as features. For each bar, the black outlined bar indicates the accuracy achieved when including the red edge bands and the colored bar shows the accuracy that was achieved when the red edge bands were not included in the classification. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

larger underestimation of 22% occurred in Schleswig-Holstein (SH) and the estimate for Brandenburg was 25% above the census area. The mapped estimate for all cereal types combined is on average 11% above the census area, but slightly underestimated for some states with large cultivation areas, namely North Rhine-Westphalia (NW) and Lower Saxony (NI) where the mapped area is 4% and 5% below the census area. The mapped maize and rapeseed classes compare very well to the census area, as indicated by low RMSE and small regression offset values (Fig. 9), with a minor tendency for overestimation of 2% and 3%, respectively. Larger underestimations for maize occurred in NI (12%) and overestimations for rapeseed in MV (10%). The comparisons for sugar beet and potato appears to be more scattered, but overall acreages are all below 75,000 ha which is much lower than for the other crops. On average, the mapped estimates are 8% below (sugar beet) and 14% above (potato) the areas reported in the agricultural census. Most notably, BY and NW are slightly underestimated for sugar beet (19% and 11%, respectively) and for potato (19% and 29%, respectively) according to our mapped acreages.

6. Discussion

Sentinel-2 and Landsat provide observations of similar nature, but differences in observing geometry, spectral bands and geolocation exist and need to be addressed. Integrating both data sources allows improving the temporal frequency of time series. Compositing provides an opportunity for integrating observations from both sensors and accordingly approaches are needed to derive intra-annual series of multi-sensor composites. Such data can greatly improve many applications focusing on dynamic land surface processes, such as agricultural monitoring and crop type mapping.

The HLS dataset is the first multi-mission, harmonized reflectance product that accounts for sensor differences in geometry, viewing angles and spectral bandpasses and minimizes these differences using state-of-the-art algorithms. This provides users with analysis-ready data that can directly be used in different application scenarios. The full normalization currently only works for 30 m data, for which S2 bands are spatially aggregated. This is unfortunate, as many regions in Europe and around the world feature small-scale farm and landscape structures and would therefore profit from higher resolution imagery. Especially agriculture in Africa and other developing countries is characterized by small scale subsistence agriculture where a 30 m resolution is clearly insufficient. One of S2 key improvements is its native spatial resolutions of 10 and 20 m and these represent an indispensable asset for agricultural monitoring in the context of food security. Combining MSI and OLI at 20 m might be a better option in the future, possibly allowing for full cross-calibration when combining 10/20 m Sentinel-2 with 30 m Landsat data. The MSI red edge bands lack corresponding bands in OLI. Neglecting these bands would be unfortunate, as they allow for improved differentiation of crops. This is why we strove to provide a simple solution using spectrally interpolated values as a proxy, but this topic requires future research to apply more sophisticated methods to data.

Our compositing approach allowed transforming the spatially and

Table 5
Confusion matrix for the point-based validation of the classification based on the 10-day interval composites (2016 features).

	Grassland	Winter cereals	Maize	Rapeseed	Summer cereals	Sugar beet	Potatoes	Grapevine	D/M forest	C forest	Built-up	Water	Sum	UAC
Grassland	795	17	9	8	19	3	4	228	9		10	2	1104	72.01
Winter cereals	42	1300	9	74	46		7	65	3	4	28	3	1581	82.23
Maize	20	27	932	3	35	45	189	22	1	3	7	3	1287	72.42
Rapeseed	1	4	2	867	11		2	1			2		890	97.42
Summer cereals	39	102	8	33	836	7	129	12	1	1	12	4	1184	70.61
Sugar beet	14	3	9	2	3	923	116				2		1072	86.10
Potatoes	9	2	7	2	14	20	488						542	90.04
Grapevine		4	4	3	2		17	621		1	5		657	94.52
D/M forest	10	4	7		7		2	4	693	75	42	17	861	80.49
C forest	2				1				249	899	18	4	1173	76.64
Built-up	42	26	11	5	18	2	46	23	12	8	798	8	999	79.88
Water	26	11	2	3	8			24	27	9	76	959	1145	83.76
Sum	1000	1500	1000	1000	1000	1000	1000	1000	995	1000	1000	1000		
PAC	79.50	86.67	93.20	86.70	83.60	92.30	48.80	62.10	69.65	89.90	79.80	95.90		
OAC	80.92													

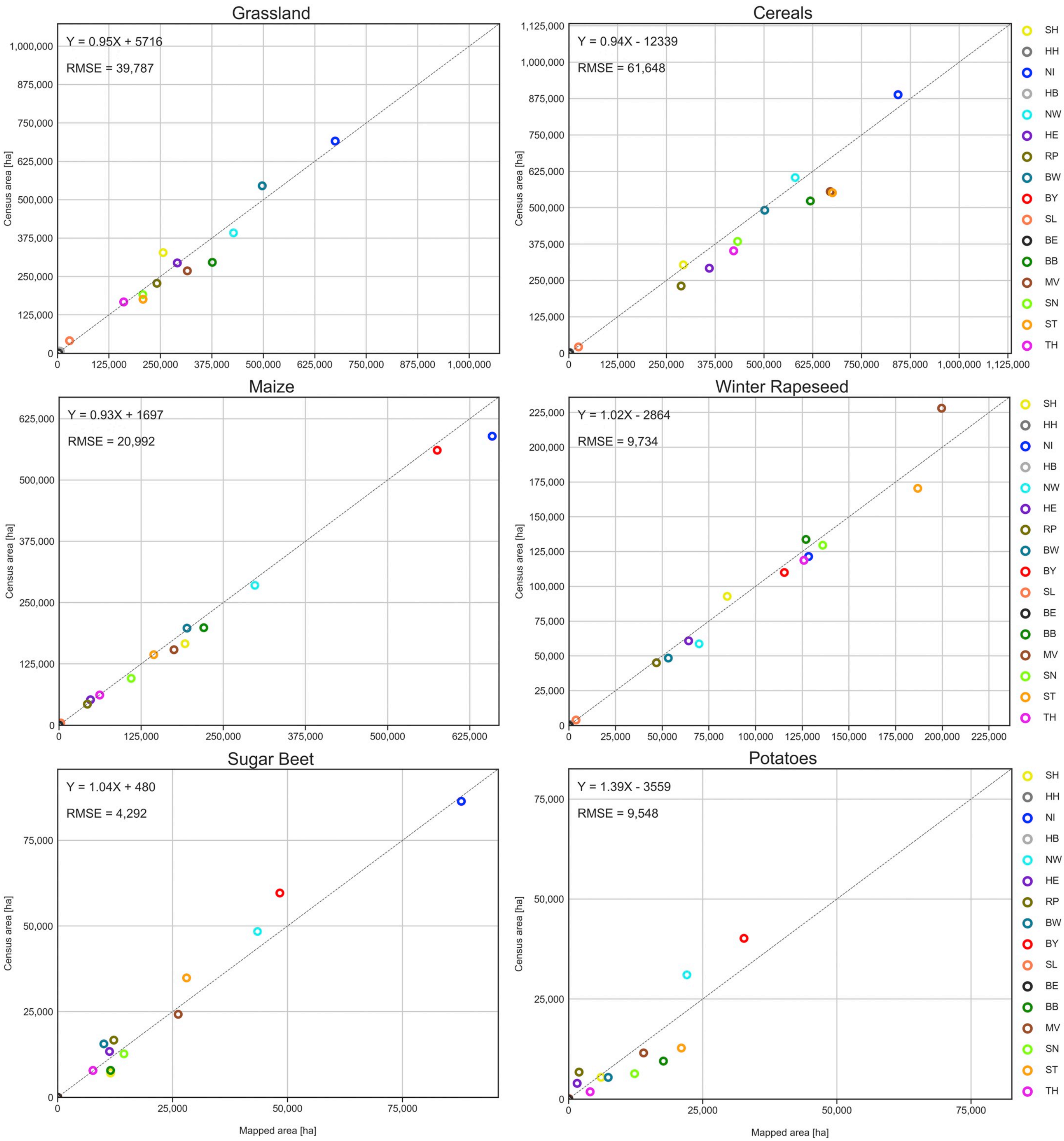

Fig. 9. Comparison of the mapped area per crop type and federal state versus the area reported in the 2016 agricultural census. Please note that small city states (BB, HH, HB) were omitted from fitting the regression.

temporally heterogeneous observations made by both sensors into equidistant reflectance composites. One key advantage of compositing is that large areas can be processed and analyzed in a systematic manner, i.e. mapping can be based on systematically generated gridded products. One disadvantage is that some cloud-free observations might not be considered. This effect is generally not that detrimental when working on narrow temporal intervals. However, similar to the above mentioned problem that the full 10 m resolution might not suffice for certain small holder landscape structures, a 10-day compositing interval is likely to be found insufficient to accurately depict specific crop developments in certain areas around the globe. Given the acquisition schedules in place for S2 and L8 during 2016 and the resulting density of observations, testing finer intervals would not have been feasible. Best-pixel compositing strategies strongly depend on high quality cloud and cloud shadow masking. While for Landsat the quality of cloud masks is already relatively high, the performance of cloud masking algorithms for S2 needs to be further optimized and the lack of TIR measurements by MSI needs to be compensated (Frantz et al., 2018). Evaluation of selected parameters such as cloud/shadow distance or HOT values is likely more suited for optimizing narrow temporal windows than selecting a best observation based on a one-dimensional measure (e.g. max NDVI, median NIR) which generally does not perform well when only few candidate observations exist for a narrow temporal window. However, our best-pixel evaluation approach is dependent on high-quality cloud screening (e.g. Fig. 6). A narrow temporal compositing interval also allows for simple temporal gap filling for which we use a linear, temporal-spectral interpolation. This approach seems largely valid as long as the temporal gap does not become too large. On the other hand, most machine learning approaches perform better even on gap-filled time series including larger temporal gaps compared to time series values that contain lots of no-data values. The *max_interval* parameter leaves this to the user decide what is most suited for a given application. Cloud coverage was especially high during the winter months, which is typical for many temperate regions, and the amount of clear sky observations is consequently greatly reduced. Identification of the most meaningful temporal periods through feature selection procedures could be used to learn more about value of the entire time series versus selected seasonal observations for mapping purposes. Our approach could additionally incorporate the spatial neighborhoods around a pixel or make use of observations acquired by SAR sensors such as Sentinel-1 (Reiche et al., 2015) or by coarse-resolution sensors such as OLCI onboard of Sentinel-3 (Donlon et al., 2012) that provides many spectral bands in the VISNIR spectral domain.

The achieved overall accuracies are promising but have to be considered with caution as the point-based validation for the agricultural classes was only performed across the three states for which the reference data was available. Our crop and land cover map achieved the highest overall and class-specific accuracies when using the full spectral features of the 10-day interval composites for training and prediction. The observed differences in class-specific accuracies between 10-day and monthly or seasonal composites were larger for highly dynamic classes such as cereal crops or rapeseed and much less for the non-agricultural classes. This finding confirms assumptions regarding the value of high temporal repeat observations for mapping dynamic phenomena such as agriculture, and that short interval composites preserve much of the required temporal information. Including the red edge bands led to improvements of overall accuracies in all cases, and had the most pronounced effect on crop classes, where several class-specific accuracies were improved. This suggests that including the red edge bands, even if those for L8 were simply derived by spectral interpolation, can lead to improvements in mapping accuracy. However, the strongest overall difference was observed for the features sets with a lower temporal resolution, suggesting that improvements at least partially depend on the type of classifier and number of features. Similar effects on the mapping accuracy were observed when we using the non-gap filled 10-day features in the RF models. While the reduction in OAC was rather gradual, class specific accuracies for several crop types were reduced considerably.

Our mapped area estimates compare very well to the census estimates, especially for cereals and maize. For grassland the agreement is slightly lower, which is not surprising considering the class heterogeneity of different grassland sub-classes in Germany. Various management schemes apply, from grasslands for conservation to high-intensity grasslands with up to five to six mowing cycles per year (Franke et al., 2012). Future iterations may include more specific crop classes (e.g. individual spring cereals, etc.) including some that were excluded from the first analysis. Many of the selected classes may still contain other crops that have not been classified due to their relatively small share in overall land cover, e.g. leafy vegetables grown in the Rhine valley being classified as sugar beet or potato. Also grapevine turned out difficult to identify, as many of the vineyards are managed with complete or alternating inter-row grass covers of very different species composition, impeding the training for this class. As our training data for vineyards was spatially restricted to Franconia in Northern Bavaria, our classification does not achieve realistic mapping results for many of the morphologically very different vineyards in the wine-growing regions along the Mosel and Rhine valleys. With further training data sets from these areas, higher accuracies are expected. Finally, it should be noted that, while this study focused only on the mapping performance using different image based datasets, the accuracy of area estimates can be further optimized by combining map-based area estimates with sample-based statistics (Gallego, 2004; Olofsson et al., 2014). Adjusting mapped crop acreages using statistics based on map validation can greatly improve area estimates and complement classification-based mapping (Kontgis et al., 2015).

National-scale assessments of crop and land cover are of great value for many subsequent analyses such as environmental impact assessments or investigating the effects of (supra-) national policies such as the EU CAP. Especially in combination with mapped agricultural management activities, crop type maps can enter into crop growth and agro-ecosystem models for simulations of yields, nitrate leaching, water loss, greenhouse gas emissions and more, towards monitoring of environmental variables that are otherwise difficult or impossible to measure across large areas. Simulation modeling on the basis of a fine-scale observed land cover map will also support the consideration of other ecosystem services in landscape management and optimization. These observed maps will gradually replace aggregated land-use approaches, in which the aggregation step is another source of uncertainty of the simulation result (Hoffmann et al., 2016), and improve large-scale assessments currently based on single-crop evaluations (Eitzinger et al., 2013; Asseng et al., 2015; Donatelli et al., 2015) by a much more elaborated crop distribution. This is especially important for assessments of long-term carbon dynamics in agricultural soils (Taghizadeh-Toosi and Olesen, 2016; Wiesmeier et al., 2016) and related emissions (Blanke et al., 2017).

7. Conclusions and outlook

The presented results show that the combined use of Sentinel-2A and Landsat data provides an improved number of observations to map the distribution of crops over large areas using a compositing processor optimized for narrow temporal intervals. This capability will be further improved by adding data from Sentinel-2B. Moreover, Sentinel-1 C-band radar data can provide weather-independent and complementary observations that could further enhance such analyses. As we found that short compositing intervals improve the mapping accuracy for most crops, future research may focus on creating a deeper understanding concerning critical phenological periods for different classification problems in different regions of the world. Crops like maize, rapeseed, sugar beet but also cereals are overall well detected and the narrow 10-day interval clearly outperforms seasonal features and leads to slightly higher accuracies compared to monthly features. The explicit derivation of phenology parameters bears great potential to further improve insight on land use and management of different classes that goes beyond conventional mapping. Here the expected increase in effective observation frequencies through Sentinel-2B will be crucial. Multi-sensor integration will commonly face problems of non-corresponding bands. Our results show, that using a simple proxy for missing red edge observations in OLI, improves mapping performance in general and specifically for crop classes. Moreover, gap-filling procedures applied to time series of 10-day interval composites further improves the machine learning based prediction of crop and land cover classes. Further research into more sophisticated approaches for gap filling or red edge proxy value generation bears great potential for time series based crop and land cover mapping. Analysis ready data such as available from HLS and higher-level gridded products, such as the 10-day composite time series, are highly valuable inputs for scaling crop type mapping to national scales and beyond.

Having produced a first single-year wall-to-wall crop map, one can soon expect multi-year analyses that disentangle crop rotation patterns using similar processing and analyses approaches. With respect to agro-ecosystem modeling, feeding observed crop rotations into simulations are expected to improve future yield and soil carbon assessments significantly, replacing the gross assumptions that still need to be made for current scenarios, and getting away from pretty unrealistic single-crop evaluations. While current, process-based agro-ecosystem simulation models, are already fit for purpose (Kollas et al., 2015), the required input data products for large-scale simulations are still at the dawn of becoming available. This also includes the simulation of grassland biomass and quality dynamics for different applied purposes. Also here, models are at hand (Kipling et al., 2016), but science still lacks appropriate methods to inform them for simulations of grassland ecosystem services. The presented intra-annual reflectance composites represent a milestone on this way towards model-data fusion for environmental monitoring and assessment. The crop and land cover map produced in this study is accessible under https://doi.pangaea.de/10.1594/PANGAEA.893195.

Acknowledgements

This research is funded through the ESA Living Fellowship Program (ESA Contract No. 4000112795/14/I-SBo) and Humboldt-Universität zu Berlin. We thank NASA for the HLS data and for including Germany as one of the test sites. We thank the Federal States of Brandenburg, Mecklenburg-Vorpommern and Bavaria for the provision of the 2016 IACS (InVeKoS) data set. This research contributes to the Landsat Science Team and the Global Land Project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2018.10.031.

References

Asseng, S., Ewert, F., Martre, P., Rotter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V.V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vile, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.K., Muller, C., Kumar, S.N., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Rezaei, E.E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, I., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce global wheat production. Nat. Clim. Chang. 5, 143–147.

Baghdadi, N., Boyer, N., Todoroff, P., El Hajj, M., Bégué, A., 2009. Potential of SAR sensors TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on Reunion Island. Remote Sens. Environ. 113, 1724–1738. https://doi.org/10.1016/j.rse.2009.04.005.

BGR, 2014. Ackerbauliches Ertragspotential der Böden in Deutschland. Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, Germany.

Blanke, J.H., Olin, S., Sturck, J., Sahlin, U., Lindeskog, M., Helming, J., Lehsten, V., 2017. Assessing the impact of changes in land-use intensity and climate on simulated trade-offs between crop yield and nitrogen leaching. Agric. Ecosyst. Environ. 239, 385–398.

Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Morin, D., Rabaute, T., Savinaud, M., Sepulcre, G., Valero, S., Defourny, P., Koetz, B., 2015. Sentinel-2 for agriculture: supporting global agriculture monitoring. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4185–4188. https://doi.org/10.1109/IGARSS.2015.7326748.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto Int. 26, 341–358.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Cihlar, J., Manak, D., Diorio, M., 1994. Evaluation of compositing algorithms for AVHRR data over land. IEEE Trans. Geosci. Remote Sens. 32, 427–437.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002.

Clewley, D., Bunting, P., Shepherd, J., Gillingham, S., Flood, N., Dymond, J., Lucas, R., Armston, J., Moghaddam, M., 2014. A python-based open source system for geographic object-based image analysis (GEOBIA) utilizing raster attribute tables. Remote Sens. 6, 6111–6135.

Davidson, A.M., Fisette, T., McNairn, H., Daneshfar, B., 2017. Detailed crop mapping using remote sensing data (Crop Data Layers). In: Delince, J. (Ed.), Handbook on Remote Sensing for Agricultural Statistics (Chapter 4). Handbook of the Global Strategy to improve Agricultural and Rural Statistics (GSARS), Rome.

Destatis, 2017. Bodennutzung der Betriebe (Landwirtschaftlich genutzte Flächen), Agrarstrukturerhebung 2016. Statistisches Bundesamt, Land- und Forstwirtschaft, Fischerei.

Donatelli, M., Srivastava, A.K., Duveiller, G., Niemeyer, S., Fumagalli, D., 2015. Climate change impact and potential adaptation strategies under alternate realizations of climate scenarios for three major crops in Europe. Environ. Res. Lett. 10.

Doninck, J.v., Tuomisto, H., 2017. Influence of compositing criterion and data availability on pixel-based Landsat TM/ETM+ image compositing over Amazonian forests. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 857–867.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.H., Femenias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., Sciarra, R., 2012. The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. Remote Sens. Environ. 120, 37–57.

Doxani, G., Vermote, E., Roger, J.-C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G., Li, F., Louis, J., Mangin, A., Pahlevan, N., Pflug, B., Vanhellemont, Q., 2018. Atmospheric correction inter-comparison exercise. Remote Sens. 10, 352. https://doi.org/10.3390/rs10020352.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25–36.

DWD, 2016. In: (DWD), G.M.S. (Eds.), Grids of Monthly Averaged Daily Air Temperature (2m) Over Germany, Version v1.0. DWD Climate Data Center (CDC), Offenbach, Germany.

Eitzinger, J., Trnka, M., Semeradova, D., Thaler, S., Svobodova, E., Hlavinka, P., Siska, B., Takac, J., Malatinska, L., Novakova, M., Dubrovsky, M., Zalud, Z., 2013. Regional climate change impacts on agricultural crop production in Central and Eastern Europe - hotspots, regional differences and common trends. J. Agric. Sci. 151, 787–812.

Flood, N., 2013. Seasonal composite Landsat TM/ETM plus images using the Medoid (a multi-dimensional median). Remote Sens. 5, 6481–6500.

Flood, N., 2017. Comparing Sentinel-2A and Landsat 7 and 8 using surface reflectance over Australia. Remote Sens. 9, 659.

Franke, J., Keuck, V., Siegert, F., 2012. Assessment of grassland use intensity by remote sensing to support conservation schemes. J. Nat. Conserv. 20, 125–134.

Frantz, D., Röder, A., Stellmes, M., Hill, J., 2017. Phenology-adaptive pixel-based compositing using optical earth observation imagery. Remote Sens. Environ. 190, 331–347.

Frantz, D., Haß, E., Uhl, A., Sto, J., Hill, J., 2018. Improvement of the Fmask Algorithm for Sentinel-2 Images: Separating Clouds From Bright Surfaces Based on Parallax Effects. https://doi.org/10.1016/j.rse.2018.04.046.

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Abdel Aziz, S., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M., 2015. Mapping global cropland and field size. Glob. Chang. Biol. 21, 1980–1992.

Gallego, F.J., 2004. Remote sensing and land cover area estimation. Int. J. Remote Sens. 25, 3019–3047.

Gallego, J., Sannier, C., Pennec, A., Dufourmont, H., 2016. Validation of Copernicus Land Monitoring Services and Area Estimation. ICASVII - Seventh International Conference on Agricultural Statistics, Rome, Italy.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44 (8), 2207–2218.

Gao, F., Masek, J.G., Wolfe, R.E., 2009. Automated registration and orthorectification package for Landsat and Landsat-like data processing. J. Appl. Remote. Sens. 3.

Gascon, F., Bouzinac, C., Thepaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Tremas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P., Fernandez, V., 2017. Copernicus Sentinel-2A calibration and products validation status. Remote Sens. 9.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Graesser, J., Ramankutty, N., 2017. Detection of cropland field parcels from Landsat imagery. Remote Sens. Environ. 201, 165–180. https://doi.org/10.1016/j.rse.2017.08.027.

Griffiths, P., van der Linden, S., Kuemmerle, T., Hostert, P., 2013. A pixel-based Landsat compositing algorithm for large area land cover mapping. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6, 2088–2101.

Hagolle, O., Kadiri, M., Morin, D., 2017. Sentinel-2 Agriculture - Detailed Processing Model for Monthly Synthesis product, Sen2Agri. Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Campbell, L.B., 2016. Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. Int. J. Digital Earth 9, 1035–1054.

Hoffmann, H., Zhao, G., Asseng, S., Bindi, M., Biernath, C., Constantin, J., Coucheney, E., Dechow, R., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Heinlein, F., Kassie, B.T., Kersebaum, K.C., Klein, C., Kuhnert, M., Lewan, E., Moriondo, M., Nendel, C., Priesack, E., Raynal, H., Roggero, P.P., Rotter, R.P., Siebert, S., Specka, X., Tao, F.L.,

Teixeira, E., Trombi, G., Wallach, D., Weihermuller, L., Yeluripati, J., Ewert, F., 2016. Impact of spatial soil and climate input data aggregation on regional yield simulations. PLoS One 11.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.
Hütt, C., Koppe, W., Miao, Y., Bareth, G., 2016. Best accuracy land use/land cover (LULC) classification to derive crop types using multitemporal, multisensor, and multi-polarization SAR satellite images. Remote Sens. https://doi.org/10.3390/rs8080684.
Inglada, J., Vincent, A., Arias, M., Marais-Sicre, C., 2016. Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. Remote Sens. https://doi.org/10.3390/rs8050362.
Johnson, D.M., Mueller, R., 2010. The 2009 cropland data layer. Photogramm. Eng. Remote. Sens. 76, 1201–1205.
Johnson, J.A., Runge, C.F., Senauer, B., Foley, J., Polasky, S., 2014. Global agriculture and carbon trade-offs. Proc. Natl. Acad. Sci. U. S. A. 111, 12342–12347.
Jönsson, P., Eklundh, L., 2004. TIMESAT - a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30, 833–845. https://doi.org/10.1016/j.cageo.2004.05.006.
Karydas, C., Gitas, I., Kuntz, S., Minakou, C., 2015. Use of LUCAS LC point database for validating country-scale land cover maps. Remote Sens. 7, 5012–5041.
Kipling, R.P., Virkajaervi, P., Breitsameter, L., Curnel, Y., De Swaef, T., Gustavsson, A.M., Hennart, S., Hoglind, M., Jaervenranta, K., Minet, J., Nendel, C., Persson, T., Picon-Cochard, C., Rolinski, S., Sandars, D.L., Scollan, N.D., Sebek, L., Seddaiu, G., Topp, C.F.E., Twardy, S., Van Middelkoop, J., Wu, L.H., Bellocchi, G., 2016. Key challenges and priorities for modelling European grasslands under climate change. Sci. Total Environ. 566, 851–864.
Kollas, C., Kersebaum, K.C., Nendel, C., Manevski, K., Muller, C., Palosuo, T., Armas-Herrera, C.M., Beaudoin, N., Bindi, M., Charfeddine, M., Conradt, T., Constantin, J., Eitzinger, J., Ewert, F., Ferrise, R., Gaiser, T., de Cortazar-Atauri, I.G., Giglio, L., Hlavinka, P., Hoffmann, H., Hoffmann, M.P., Launay, M., Manderscheid, R., Mary, B., Mirschel, W., 2015
Kontgis, C., Schneider, A., Ozdogan, M., 2015. Mapping rice paddy extent and intensification in the Vietnamese Mekong River Delta with dense time stacks of Landsat data. Remote Sens. Environ. 169, 255–269. https://doi.org/10.1016/j.rse.2015.08.004.
Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14 (5), 778–782.
Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Lück, W., van Niekerk, A., 2016. Evaluation of a rule-based compositing technique for Landsat-5 TM and Landsat-7 ETM+ images. Int. J. Appl. Earth Obs. Geoinf. 47, 1–14.
Maus, V., Camara, G., Cartaxo, R., Sanchez, A., Ramos, F.M., de Queiroz, G.R., 2016. A time-weighted dynamic time warping method for land-use and land-cover mapping. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 3729–3739.
McNairn, H., Kross, A., Lapen, D., Caves, R., Shang, J., 2014. Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2. Int. J. Appl. Earth Obs. Geoinf. 28, 252–259. https://doi.org/10.1016/j.jag.2013.12.015.
Michael, C., David, T., 2017. Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice. Environ. Res. Lett. 12, 064016.
Mueller, L., Schindler, U., Shepherd, T.G., Ball, B.C., Smolentseva, E., Hu, C., Hennings, V., Schad, P., Rogasik, J., Zeitz, J., Schlindwein, S.L., Behrendt, A., Helming, K., Eulenstein, F., 2012. A framework for assessing agricultural soil quality on a global scale. Arch. Agron. Soil Sci. 58, S76–S82.
Nagol, J.R., Sexton, J.O., Kim, D.-H., Anand, A., Morton, D., Vermote, E., Townshend, J.R., 2015. Bidirectional effects in Landsat reflectance estimates: is there a problem to solve? ISPRS J. Photogramm. Remote Sens. 103, 129–135.
Nelson, K.J., Steinwand, D., 2015. A Landsat data tiling and compositing approach optimized for change detection in the conterminous United States. Photogramm. Eng. Remote. Sens. 81, 573–586.
Nidamanuri, R.R., Zbell, B., 2011. Use of field reflectance data for crop mapping using airborne hyperspectral image. ISPRS J. Photogramm. Remote Sens. 66, 683–691.
Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015.
Ozdogan, M., Woodcock, C.E., 2006. Resolution dependent errors in remote sensing of cultivated areas. Remote Sens. Environ. 103, 203–217.
Palmieri, A., Dominici, P., Kasanko, A., Martino, L., 2011. Diversified Landscape Structure in the EU Member States - Landscape Indicators From the LUCAS 2009 Survey, Statistics in Focus. EUROSTAT - European Comission.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Potapov, P.V., Turubanova, S.A., Hansen, M.C., Adusei, B., Broich, M., Altstatt, A., Mane, L., Justice, C.O., 2012. Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ data. Remote Sens. Environ. 122, 106–116. https://doi.org/10.1016/j.rse.2011.08.027.
Reiche, J., Verbesselt, J., Hoekman, D., Herold, M., 2015. Fusing Landsat and SAR time series to detect deforestation in the tropics. Remote Sens. Environ. 156 (0), 276–293. https://doi.org/10.1016/j.rse.2014.10.001.
Rockström, J., Williams, J., Daily, G., Noble, A., Matthews, N., Gordon, L., Wetterstrand, H., Declerck, F., Shah, M., Steduto, P., de Fraiture, C., Hatibu, N., Unver, O., Bird, J., Sibanda, L., Smith, J., 2017. Sustainable intensification of agriculture for human prosperity and global sustainability. Ambio 46, 4–17.
Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Loveland, T.R., Vermote, E., Zhang, C., 2010. Web-enabled Landsat Data (WELD): Landsat ETM plus composited mosaics of the conterminous United States. Remote Sens. Environ. 114, 35–49.
Roy, D.P., Zhang, H.K., Ju, J., Gomez-Dans, J.L., Lewis, P.E., Schaaf, C.B., Sun, Q., Li, J., Huang, H., Kovalskyy, V., 2016. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. Remote Sens. Environ. 176, 255–271.
Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., Zhang, L., 2015. Missing information reconstruction of remote sensing data: a technical review. IEEE Geosci. Remote Sens. Mag. 3 (3), 61–85.
Storey, J.C., Roy, D.P., Masek, J., Gascon, F., Dwyer, J.L., Choate, M.J., 2016. A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. Remote Sens. Environ. 186, 121–122.
Taghizadeh-Toosi, A., Olesen, J.E., 2016. Modelling soil organic carbon in Danish agricultural soils suggests low potential for future carbon sequestration. Agric. Syst. 145, 83–89.
Testa, S., Soudani, K., Boschetti, L., Borgogno Mondino, E., 2018. MODIS-derived EVI, NDVI and WDRVI time series to estimate phenological metrics in French deciduous forests. Int. J. Appl. Earth Obs. Geoinf. 64, 132–144. https://doi.org/10.1016/j.jag.2017.08.006.
Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. U. S. A. 108, 20260–20264.
Tóth, K., Kučas, A., 2016. Spatial information in European agricultural data management. Requirements and interoperability supported by a domain model. Land Use Policy 57, 64–79. https://doi.org/10.1016/j.landusepol.2016.05.023.
Udelhoven, T., 2011. TimeStats: a software tool for the retrieval of temporal patterns from global satellite archives. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 4, 310–317. https://doi.org/10.1109/JSTARS.2010.2051942.
Vancutsem, C., Pekel, J.F., Bogaert, P., Defourny, P., 2007. Mean compositing, an alternative strategy for producing temporal syntheses. Concepts and performance assessment for SPOT VEGETATION time series. Int. J. Remote Sens. 28, 5123–5141.
Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56.
Vuolo, F., Ng, W.-T., Atzberger, C., 2017. Smoothing and gap-filling of high resolution multi-spectral time series: example of Landsat data. Int. J. Appl. Earth Obs. Geoinf. 57, 202–213. https://doi.org/10.1016/j.jag.2016.12.012.
White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., Coops, N.C., Hall, R.J., Hostert, P., Dyk, A., Guindon, L., 2014. Pixel-based image compositing for large-area dense time series applications and science. Can. J. Remote. Sens. 40, 192–212.
Wiesmeier, M., Poeplau, C., Sierra, C.A., Maier, H., Fruhauf, C., Hubner, R., Kuhnel, A., Sporlein, P., Geuss, U., Hangen, E., Schilling, B., von Lutzow, M., Kogel-Knabner, I., 2016. Projected loss of soil organic carbon in temperate agricultural soils in the 21st century: effects of climate change and carbon input trends. Sci. Rep. 6.
Wolfe, R.E., Roy, D.P., Vermote, E., 1998. MODIS land data storage, gridding, and compositing methodology: level 2 grid. IEEE Trans. Geosci. Remote Sens. 36, 1324–1338.
Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., Landsat Sci, T., 2008. Free access to Landsat imagery. Science 320, 1011.
Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ. 122, 2–10.
Wulder, M.A., Hilker, T., White, J.C., Coops, N.C., Masek, J.G., Pflugmacher, D., Crevier, Y., 2015. Virtual constellations for global terrestrial monitoring. Remote Sens. Environ. 170, 62–76.
Yan, L., Roy, D., Zhang, H., Li, J., Huang, H., 2016. An automated approach for sub-pixel registration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery chyba tak DO PREZENTACJI. Remote Sens. 8. https://doi.org/10.3390/rs8060520.
Zhang, Y., Guindon, B., Cihlar, J., 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. Remote Sens. Environ. 82, 173–187.
Zhu, Z., 2017. Change detection using landsat time series: a review of frequencies, preprocessing, algorithms, and applications. ISPRS J. Photogramm. Remote Sens. 130, 370–384.
Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94.
Zhu, Z., Woodcock, C.E., Holden, C., Yang, Z., 2015. Generating synthetic Landsat images based on all available Landsat data: predicting Landsat surface reflectance at any given time. Remote Sens. Environ. 162, 67–83.